

10Z AIGGACGGHACCGGITHAICGATTAICCAACCAACCGACIICCAACCGIACGAGGA 201

17 pAsnSerProProgiuProSerProgiuSerLeuAArgHisValSera 34
 232 CAATTCCTACCGCCTGAGCCATACCGAGAGTCTCCGTCATATGTAAGCC 281
 34 rGMeLLeasnSerAsnHisTyrThrSerProSerArgThrIleTyrSer 50
 282 GTATATGACACAGACCACTTACACCTTCTCGAACAATCTACTGCC 331
 51 AspargPheIleProSerArgSerAlaSerLysPheAlaLeuPheAsp11 67
 332 GATAGGTCATCCGAGTAGATGCTCTCGAAATTCGCTTGTGTTGATAT 381
 67 eAsnThrProThrGluArgAspAspSerSerSerAlaTyrThrThrL 84
 382 CAATTCCTCCAGACAGACCGCATGATGATTCACGCGCTTATACGACTC 431
 84 euLeuArgThrAlaLeuPheGlyProAspValAlaGlyProValThrPro 100
 432 TTCGTGAGAACGGCGTGTGTGGACCGGATGTGGCGGTCCGGTTACGGCG 481
 101 GluLysThrAspSerProSerMetThrLeuProAsnArgAsnIlePheArg 117
 482 GAAAAAACCGACTCCGCTCGATGACATGCGCAATAGAAATATTTTAC 531
 117 GTTCTGThrGluThrArgGlnSerMetHisSerLeuSerProPheMetA 134
 532 GTATTAAGACGAGACGACAGACATCATGCTCCGCTTCCGCGCTTATAG 581
 134 sPAspAspPheValProGlyValAsnHisSerProValLysAlaProArg 150
 582 ATGATGATTTGTTCTCTGTTATCATCTCCGTTAAAGCTCCACAGG 631
 151 LysValProArgSerProTyrLysValLeuAspAlaProAlaLeuGlnAs 167
 632 AAGGTTCTCGATCGCTTATTAAGGTTTGGATGACACTGCTTGGCAGA 681
 167 pAspPheTyrLeuAsnLeuValAspTyrSerSerHisAsnValLeuAlaV 184
 682 TGATTTTATCTGATCTGTAGATGTGTCTTACACAAATGTGTGGCTG 731
 184 alGlyLeuGlyAsnCysValTyrLeuTyrPAsnAlaCysSerSerLysVal 200
 732 TTGGTTTGGTAACTGCTCTATCTCTGGAATGCTTGGACGACAGGTA 781
 201 ThrLysLeuCysAspLeuGlyValAspAspCysValCysSerValGlyTyr 217
 782 ACTAAATATGATTTGGGGGTGATGATGTGTGTCTGTGTGGTTG 831
 217 pAlaGlnArgGlyThrHisLeuAlaValAlaGlyThrAsnAsnGlyLysValG 234
 832 GGCTCAACGTGCTACTCATCTTGGTGGAACTAACAAATGTAAGTTC 881
 234 InIleTyrAspAlaAlaArgCysLysLysIleArgSerMetGluGlyHis 250
 882 AGATTTGGGATGCACAGAGATGCCAAGATTAACATCAATGAGCGCCAT 931
 251 ArgLeuArgValAlaLeuAlaTyrSerSerSerLeuLeuSerSerG1 267
 932 CGGTTACTGTGCGGCTTGGCGCTGGAGTTCATCTCTTTGCTCTCGG 981
 267 yGlyArgAspLysAsnIleTyrGlnArgAspIleArgThrGlnGluAsp 284
 982 TGGACGGATTAAGATATTATCAACGAGATATACGACACACAGAGATT 1031
 284 heValSerLysLeuSerGlyHisLysSerGluValCysGlyLeuLysTyr 300
 1032 TTGTTAGTAACTGTCCAGACACAAATCAGAGGTTTGGAGCTGAGGTG 1081
 301 SerTyrAspAsnArgGluLeuAlaSerGlyLysAsnAspAsnLysLeuPh 317
 1082 TCATATGATTAACCGTAGGTGCTGATCTGAGAGAAATGACAACAATTTGTT 1131

317 eValTyrPAsnGlnHisSerThrGlnProValLeuLysTyrCysGluHisT 334
 1132 TGTTTGGAAATCAACACTCAACCCAGCTGTCTCTCAAGTACTGTGAGACA 1181
 334 hrAlaAlaValLysAlaIleAlaTyrSerProHisLeuHisGlyLeuLeu 350
 1182 CACACACTGTAAAGCTATGTGATGCTCTCATCTTCATCTGACTTCTT 1231
 351 AlaSerGlyGlyThrAlaAspArgCysIleArgPheThrAsnThrTh 367
 1232 GCATCTGAGAGAGAACTGACATAGATGATTTCTTTTGGAAATACAC 1281
 367 rThrAsnSerHisLeuSerCysMetAspThrGlySerGlnValCysAsnL 384
 1282 CACAACTCCACCTTAGCTAGCTGATGACACTGGAATCAGGTTTGCAATC 1331
 384 euValTyrSerLysAsnValAsnGluLeuValSerThrHisGlyTyrSer 400
 1332 TTGCTGTGTCACAAATGTCAACGAACTAGTAAGCACATGGGTACTCC 1381
 401 GlnAsnGlnIleLeValTyrPargTyrProThrMetSerLysLeuAlaTh 417
 1382 CAGAACCAATTAATGTGTGGAGATACCCCATATCTCAAAAGCTGGCAGC 1431
 417 rLeuThrGlyHisThrTyrArgValLeuTyrLeuAlaIleSerProAspG 434
 1432 TCTTACCGGCATACCTATAGGGTCTCTATCTTGCATCTCCAGATG 1481
 434 LysGlnThrIleValIleThrGlyAlaGlyAspGluThrLeuArgPheThrAsn 450
 1482 GACAGACTATGTAACTGAGCTGAGAGTAAGAACGTTAGGTTCTGGAAAT 1531
 451 ValPheProSerProLysSerGlnAsnThrGluSerGluIleGlyAlaLe 467
 1532 GTTTTCCCTCCCTTAATCAGAGATCTGAAGAGTAATGGAGCAATT 1581
 467 uSerLeuGlyArgThrThrIleArg 475
 1582 ATCTCTGGAGAACTACTATACAG 1606
 seq_name: gp_pat:AX008931

seq_documentation_block:
 LOCUS AX008931 2006 bp DNA linear PAT 06-SEP-2000
 DEFINITION Sequence 1 from Patent WO9964451.
 ACCESSION AX008931
 VERSION AX008931.1 GI:9996328
 KEYWORDS
 SOURCE
 ORGANISM
 Medicago sativa.
 Medicago sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.
 1 (bases 1 to 2006)
 Cebolla,A., Kondorosi,E. and Kondorosi,A.
 Plant protein with repeated wd40 motifs, nucleic acid coding for
 said protein, and uses thereof
 Patent: WO 9964451-A 1 16-DEC-1999;
 CEBOLLA ANGEL (FR); KONDOROSI EVA (FR); KONDOROSI ADAM (FR); CENTRE.
 NAT RECH SCIENT (FR)

FEATURES
 source
 1..2006
 Location/Qualifiers
 /organism="Medicago sativa"
 /db_xref="taxon:3879"
 182..1609
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC07447.1"
 /db_xref="GI:9996329"
 /translation="MDGNGNLPPTSTVDNSPPPEPSPESLRHYSRMINSNHTSP
 SRTIYSDFITRSASKFALFDITPTPEGSDSDSSATYTLRLALGCPDVAGVTPBEK
 TDSPSMTLPNNIRFRTETKQSNHSLSPFNDDDFYGVNHSFVKAPRKVPKSPYIVL

1082 TCATATGATTAACCGTAGGTGCTGATCTGAGAGAAATGACAACAATTTGTT 1131

DABADODEFYTLNLVDMSHNVALVGLGNCVYLMMNCSKVKTKLCLDGLVDDCVCSVGMV
 ORGTFLAVGTNNKQVQIMDAARCKIRSMEGRLRGLALMSSLSLSSGGRDKNTYOR
 DRTQEDPVSKNSGKSEVCKGLKMSYDNRELASGGNDKLFWNHOHSITQPVLYKCEHT
 AAKKAIANSPIHLGLIASGGTADRCIRRMNTTNSHLSMCTGSGOYCNLWMSKRVNE
 LYSTHYSOMQITVWRYTPMSKIAITITGHTRVLYLAISPDQITVTGAGDETLRFWN
 VFPSPKSNPTESIGLSLGRRTIR
 BASE COUNT 555 a 398 c 441 g 612 t
 ORIGIN

alignment_scores:

Quality: 2555.00 Length: 475
 Ratio: 5.379 Gaps: 0
 Percent similarity: 100.000 Percent identity: 100.000

alignment_block:

US-09-701-572-2 x AX008931 ..

Align seg 1/1 to: AX008931 from: 1 to: 2006

1 MetAspGlyThrGlyAsnArgAsnProProProThrSerThrValArgAs 17
 |||||
 182 ATGAGCGAGACCGGTAATCGAATCCACCGACTTCCACCGTCAGAGA 231
 |||||
 17 PAsnSerProProProGluProSerProGluSerLeuArgHisValSerA 34
 |||||
 232 CAATTCTCCACCGCCTGAGCCATCCCGAGAGAGTCCCTCATGTAAAGCC 281
 |||||
 34 rgmEtlAsnSerAsnHisThrThrSerProSerArgThrIleTyrSer 50
 |||||
 282 GTATGATCAACAGCAACCATTCACCTCCTCCGAAACATCTACCTCC 331
 |||||
 51 AspArgPheIleProSerArgSerAlaSerLysPheAlaLeuPheAspI 67
 |||||
 332 GATAGGTTCAATCCGAGTAGATGCTGCTCGAAATCGCTTGTGTGATAT 381
 |||||
 67 eAsnThrProThrGluArgAspAspSerSerSerAlaTyrThrThr 84
 |||||
 382 CAATCTCCGACAGAGAGCGCATGATGATGTCAGCGCTTATTCAGACTC 431
 |||||
 84 euleuArgThrAlaLeuPheGlyProAspValAlaGlyProValThrPro 100
 |||||
 432 TTCTGAGAACGGCGTTGTTGGACCGGATGTGCGCGTCCGTTACGCGC 481
 |||||
 101 GluIystrAspSerProSerMetThrLeuProAsnArgAsnIlePheAr 117
 |||||
 482 GAAAAAACCGACTCGCGTCGATGACATTCGCCAATAGGATATTTTAG 531
 |||||
 117 gTyrIystrThrGluThrArgGlnSerMetHisSerLeuSerProPheMe 134
 |||||
 532 GTATTAAGACGGAGACAGACAGTCATGCTCGCTTCGCCGTTTAGG 581
 |||||
 134 sPAspAspPheValProGlyValAsnHisSerProValIysAlaProArg 150
 |||||
 582 ATGATGATTTTGTCTCTGTTAATCATAGTCCGTTAAGGCTCTTAGG 631
 |||||
 151 IysValProArgSerProTyrIysValLeuAspAlaProAlaLeuGlnAs 167
 |||||
 632 AAGGTTCTCGATCGCTTATTAAGGTTTGGATGACACTGCTTTCAGAGA 681
 |||||
 167 pAspPheThrLeuAsnLeuValAspTrpSerSerHisAsnValLeuAla 184
 |||||
 682 TGATTTTATCTGAATCTGAGATGATGCTTCACACAAATGTGTGCTG 731
 |||||
 184 aIGlyLeuGlyAsnCyValTyrLeuTrpAsnAlaCySerSerIysVal 200
 |||||
 732 TTGGTTTGGTAACTGCTCTATCTCTGGAATGCTGTAGCAGCAAGTA 781
 |||||
 201 ThrIysLeuCyAspLeuGlyValAspAspCyValCySerValGlyTr 217
 |||||
 782 ACTAATATATGATTTGGGGTGTGATGATGCTTGTCTGTTGTTGCTG 831
 |||||
 217 palaaGlnArgGlyThrHisLeuAlaValGlyThrAsnAsnIyLysValG 234

|||||
 832 GGCCTCAACCGGTACTCATCTTCTCTTGGAACTAACATGTGAAAGTTC 881
 |||||
 234 InIleTrpAspAlaIleArgCyValLysIleLeuArgSerMetGluGlyHis 250
 |||||
 882 AGATTGGGATGCGCAAGATGCGAAGATTAAGATCAATAGGAGGCCAT 931
 |||||
 251 ArgLeuArgValGlyAlaLeuAlaIleTrpSerSerSerLeuLeuSerG 267
 |||||
 932 CGGTTACGTGTCGGGGCTTGCGCTGGAGTTCATCTTGTGCTCTG 981
 |||||
 267 yGlyArgAspLysAsnIleTyrGlnArgAspIleArgThrGlnGluAsp 284
 |||||
 982 TGGACGGGATAGAAATATTATCAACAGATATATACACACAGAGAGATT 1031
 |||||
 284 heValSerIysLeuSerGlyHisIysSerGluValCyGlyLeuIystrp 300
 |||||
 1032 TTGTTACTTAACGTCTCAGACACAAATCAGAGGTTTGTGACTGAAGTGG 1081
 |||||
 301 SerTyrAspAsnArgGluLeuAlaSerGlyGlyAsnAspAsnLysLeuPh 317
 |||||
 1082 TCATATGATTAACCGTGAAGTGGCATCTGGAGGAAATGACAAATTTGTT 1131
 |||||
 317 eValTrpAsnGlnHisSerThrGlnProValLeuIystrYrCyGlnHis 334
 |||||
 1132 TGTTTGAATCAACACTCAACCCAGCCTGTCTCAAGTACTGTGACACCA 1181
 |||||
 334 hrAlaAlaValIysAlaIleAlaTrpSerProHisLeuHisGlyLeuLeu 350
 |||||
 1182 CAGCAGCTGTTAAAGCTATTGCAATGCTCTCTCATCTTCATGACCTTCT 1231
 |||||
 1232 GCATCTGAGAGAGAACTCGAGATAGATGATGCTTTTGGAAATCAAC 1281
 |||||
 367 rThrAsnSerHisLeuSerCyMetAspThrGlnIystrGlnValCyAsnL 384
 |||||
 1282 CACAAACTCAACCTTAGCTGTATGAGACACTGGAAGTCAGAGTTTGCAAATC 1331
 |||||
 384 euValTrpSerIysAsnValAsnGluLeuValSerThrHisGlyTyrSer 400
 |||||
 1332 TTGCTGTCGTCAAAATATGCAAGCACTAGTAAAGCAACATGGGTACTCC 1381
 |||||
 401 GlnAsnGlnIleIleValIleTrpArgTyrProThrMetSerIysLeuAla 417
 |||||
 1382 CAGAACCGAGATTATTTGAGATATCCCACTATGTCAAAAGCTGGCGAC 1431
 |||||
 417 rLeuThrGlnHisThrTyrArgValLeuTyrLeuAlaIleSerProAsp 434
 |||||
 1432 TCTTACCGGCATCTTATAGGGTGTCTATCTTCCATCTTCCAGATG 1481
 |||||
 434 LysGlnThrIleValThrGlyAlaGlyAspGluThrLeuArgPheTrpAsn 450
 |||||
 1482 GACAGACTATTGTAACTGAGCGCTGGAGATGAAGCGCTAGGTTCTGGAAT 1531
 |||||
 451 ValPheProSerProLysSerGlnAsnThrGluSerGluIleGlyAlaLe 467
 |||||
 1532 GTTTTCCCTTCCCTTAATCAAGAAATACAGAAAGTGAAGATTCGAGCAATT 1581
 |||||
 467 uSerLeuGlyArgThrThrIleArg 475
 |||||
 1582 ATCTCTTGGAGAACTACTATTCAGG 1606

seq_name: gb_pl:AY063875

seq_documentation_block:

LOCUS AY063875 1911 bp mRNA linear PLN 05-DEC-2001
 DEFINITION Arabidopsis thaliana putative Strv1 protein (At4g11920) mRNA,
 complete cds.
 ACCESSION AY063875
 VERSION AY063875.1 GI:17380837
 KEYWORDS FLI CDNA.
 SOURCE thale cress.

ORGANISM

Arabidopsis thaliana

REFERENCE

Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta; Spermatophyta; Magnoliophyta: eudicotyledons: core eudicots; Rosidae: eurosids II: Brassicales: Brassicaceae: Arabidopsis.

AUTHORS

1 (bases 1 to 1911)

TITLE

Full length cDNA of gene At4g11920 (GI:15234350)

JOURNAL

2 (bases 1 to 1911)

REFERENCE

Yamada, K., Ban, J., Banno, F., Chang, E., Dale, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Yu, S., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Koeseima, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shin, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.

COMMENT

Submitted (19-NOV-2001) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA

TITLE

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPL cDNAs (RAPL cDNA: RIKEN Arabidopsis Full-length cDNA): Seki, M., Narusaka, M., Ishida, J., Hayashizaki, Y., and Shinozaki, K.

The Salk, Stanford, PEGC (SSP) Consortium members carried out the sequencing and annotation of the RAPL cDNAs: Yamada, K., Ban, J., Banno, F., Chang, E., Dale, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Yu, S., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Koeseima, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shin, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PEGC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PEGC) contributed equally to this work as PIs.

FEATURES

Location/Qualifiers

1..1911

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

/chromosome="4"

/clone="RAF106-86-G12 (R12065)"

/note="This clone is in a modified pBluescript vector (lambda ZAP) as a XhoI/SstI insert."

ecotype: Columbia"

1..1911

/gene="At4g11920"

1..125

/gene="At4g11920"

126..1553

/gene="At4g11920"

/codon_start=1

/evidence="experimental"

/product="putative Swi1 protein"

/protein_id="FAL36231.1"

/db_xref="GI:17380838"

/translation="MEEDSTTPKKKSDQLNPSPMNPPTYSLESRINRLIDSNHY

KSPKPIYSDFIPSRGSGNFALFLASSPNNKDGEDAGSYASLTKTALFGPTPE

KSDVNGFSPSGNIFRFTKSTQKSLNIPSPDVYGVSPSPVSRKILKSPKVL

DAPALDDFFLNLVDMSAONVLAAGLNCVYLMAACSSVTKICDGLGVETCVGWA

LRGTHLAIGTSSTVOIMVLRCKNIRMEGRIRLGAALMASSVSGSRKSIOR

DIRTFQDHYKSLKGRKSGTICGLKSSDNEELASGDNKLPFWMNOSHSTPVRFEHA

AAVAKTAMSPHFHGLASGGGADRCIRFMNTYTNHLCLCVDTNSKLVMSKVN

LVSTHGYSONQIIVWKYPTMSKIALTLTGHSRYLVLAASPDQITVYTGADFLRW

3'UTR 1554..1911
/gene="At4g11920"

misc_difference 1894
/gene="At4g11920"
/note="compared to genomic sequence"
/replace="L"

BASE COUNT 577 a 349 c 450 g 535 t
ORIGIN

alignment_scores:
Quality: 1908.00 Length: 483
Ratio: 4.458 Gaps: 6
Percent Similarity: 88.613 Percent Identity: 75.362

alignment_block:
US-09-701-572-2 x AY063875 ..

Align seg 1/1 to: AY063875 from: 1 to: 1911

1 MetAspGlyThrGlyAsnArgAsnProProProThSerThValArgAs 17

126 ATGGAAGAGATGATCAACACACGAGAAAGAAAGATTCAGCTCAGCT 175

17 pAsnSerProPro.....ProGluProSerProGluSerLeuA 30

176 GAATCTTCACCGCTCGATGAATGCTCCGACGCTCAGCTCAGCTCA... 221

30 rghIsValSerArgMetIleAsnSerAsnHisTyrThSerProSerArg 46

222 ..GGAATCAACCGCTTAAATGATGATCAATCATCATCATCTCTTGGAA 269

47 ThrIleTyrSerAspArgPheIleProSerArgSerAlaSerIlePhe 63

270 CCTATTACTCAGATAGGTTTATTCAGATGATCTGCTTCATTTCCG 319

63 aLeuPheAspIle.....AsnThrProThr.....GluGlyArgAspA 76

320 GCCTTTCGATTTGATGATCTCTTCGCGGAAATAGAAAGATGGAAAG 369

76 spSerSerSerAlaTyrThrThrLeuLeuArgThrAlaLeuPheGlyPro 92

370 ATGGCGGTGCTTCTTATGCGAGCTTTGAAACGCGCTTTT..... 413

93 AspAlaAlaGlyProValThrProGluTyrThrAspSerProSerMet 109

414GCTCCGGTGGACCGCGGAGAAAGATGATCTTAAATGGCT 454

109 rLeuProAsnArgAsnIlePheArgTyrIleThrGluThrArgGlnSer 126

455 TTCCTCGCTCGGGAATATTTTGTAGTTAAAGCAAGCAAGCAAGCTTT 504

126 ethIsSerLeuSerProPheMetAspAspAspPheValProGlyValAsn 142

505 TGAATTTGATTCGCGCTTTT...GATTCGATGCTGTTGCTGCTTAA 551

143 HisSerProValLysAlaProArgLysValProArgSerProTyrIle 159

552 CCTAGTCTCTTAAAGTCGCGGAGAAAGATTTCTTAGTCTCTTATAGT 601

602 ATTGATGCGCGCGCTCTGCAAGATGATTTTCTTGAATCTCTGATG 651

176 rPserSerHisAsnValLeuAlaValGlyLeuGlyLysGlyValTyrLeu 192

652 GGTTCGGCAACAAATGTTCTTGCAGTGGGATTAAGCAAGCGCTGATTT 701

193 TrpAsnAlaLysSerSerLysValThrIleLysLeuCysAspLeuGlyVal 209

702 TGGAAATGCTTGTAGTACAGAGTAACTTAAGTTATGCGATCTTGCG 751

209 PASPCysValLysSerValGlyTPRAlaGlnArgGlyThrHisLeuAlav 226
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
752 TGAACCTGTGGCTCAGTTGGTGGGCACACTACGTGAACACATTTGGCTA 801

226 ALGlyThrAsnAsnGlyLysValGlnIleTPRASPalaAlaArgCysLys 242
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
802 TTGGAACAGTAGCGGACAGCAAGTAAGATATGGAGTGCTTACGATGCAAG 851

243 LysIleArgSerMetGluGlyHisArgLeuArgValIGlyAlaLeuAlaTr 259
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
852 AATAATTAAACAATGCAAGGTCATCGGTAAGAAGTTGAAGCCCTGGCATG 901

259 pSerSerSerLeuLeuSerSerGlyIARqspLysAsnIleTryGlna 276
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
902 GAGCTCATCTGTCTGTCTTCAGTAGTAGAGACAAGAGCAATCTTCAGA 951

276 rGaSPileArgThrGlnGluAspPheValSerLysLeuSerGlyHisLys 292
952 GAGACATTCGTAACTCACAGAGATCATGTCAGTAAACCTTAAAGTGCACAAA 1001

293 SerGluValCysGlyLeuLysTrySerTryAspAsnArgGluLeuAlaSe 309
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1002 TTCGAAATCTGTGGACTCAATGTGCTTCGGCAATTCGGAGACTAGCATC 1051

309 rGLyGlyAsnAspAsnLysLeuPheValITPRasnGlnHisSerThrgInp 326
1052 AGGTGGAACGCACATTAAGCTGTTGTGTGGAACCAACATTCACAAACAAC 1101

326 rovalLeuLysTryCysGlnHisThrAlaAlaValLysAlaIleLeuATRP 342
1102 CGGTTTTGAGATTCTGTGAACACGCCGCGCACGTGAACCCCATGTGGTGG 1151

343 SerProHisLeuHisGlyLeuLeuAlaSerGlyGlyGlyThrAlaAspAr 359
1152 TCCTCCACATCATTTTGGACTCTCTGCTCTGTGGTGCGCACGTGATAG 1201

359 gCysIleArgPheTrpAsnThrThrThrasSerHisLeuSerCysMeta 376
1202 ATGTATCCGTTTCTGTGAACACACACAAACACACTCTTTAAATTTGGGTAG 1251

376 sPTHisSerGlnValCysAsnLeuValITPRSerLysAsnValAsnGlu 392
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1252 ATACCACACGCCAGGTGTGTATTTGGTATGTGCTTAAGCAAGCGAATGAG 1301

393 LeuValSerThrHisGlyTrySerGlnAsnGlnIleIleValITPRArgTy 409
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1302 CTTGTGAGACACACACGGGTATTTCCAGAACCAATCATGTGTGGAAAAA 1351

409 rProThreSerLysLeuAlaThrLeuThrgLysHisThrTryArgValI 426
1352 TCCAAACCATGTATAATTGGCAACCTCTACGTGTCACTGTACCGGGTTC 1401

426 euTYrLeuAlaIleSerProAspGlyGlnThrIleValThrgLyslaGly 442
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1402 TGACCTTCGGGTGCACCGGATGGACACAAATTTGTACAGAGCAGAGA 1451

443 AspGluThrLeuArgPheTrpAsnValPheProSerProLysSerGlnAs 459
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1452 GATGAACCTTGAATCTGGAATGTCTTCCCTTCCTCCGAAATCTAGAG 1501

459 nThrgLysSerGluIleGlyAlaLeuSerLeuGlnArgTyThrgThrIleArg 475
: |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1502 CAGGAGACGCAAAATCGGGCATTTATCTTTGGTAGAACACAAATCCGG 1550

seq_name: gb_pl:AF134835

seq_documentation_block:

LOCUS AF134835 8450 bp DNA linear PLN 05-MAR-2000

DEFINITION Medicago truncatula WD-repeat cell cycle regulatory protein (CCS52)

FEATURES

Accession AF134835 GI:7158291

KEYWORDS

Source	Organism	Reference Authors	Title	Journal	Features
barrel medic.	Medicago truncatula	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.			
1 (bases 1 to 8450)	Cebolla,A., Vinardell,J.M., Olah,B., Kiss,E., Roudier,F., Kondorosi,A. and Kondorosi,E.		The mitotic inhibitor cc52 is required for endoreduplication and polyploidy-dependent cell enlargement in plants		
2 (bases 1 to 8450)	Vinardell,J.M. and Kondorosi,E.		Unpublished		
Submitted (12-MAR-1999)	ISV, CNRS, Avenue de la Terrasse, Bat. 23, Gif Sur Yvette 91198, France				
Location/Qualifiers					
1. 8450	Location/Qualifiers				
/organism="Medicago truncatula"					
/strain="Ghor"					
/db_xref="taxon:3880"					
join<2461..2934,3022..3144,3295..3399,5309..5488,5512..5634,5735..5929,6459..6560,6651..6713,6827..6895,7452..>7505					
/gene="CCS52"					
/product="CCS52"					
<2461..>7505					
/gene="CCS52"					
join(2461..2934,3022..3144,3295..3399,5309..5488,5512..5634,5735..5929,6459..6560,6651..6713,6827..6895,7452..7505)					
/gene="CCS52"					
/codon_start=1					
/product="WD-repeat cell cycle regulatory protein"					
/db_xref="GI:7158292"					
/protein_id="AF37386.1"					
/translation="MDGTGRNRPPTSTVGDNPPPEPSPESLRHVSRMINSNHTSPSRITYSRFRIPRSASKEFALPDINPTBERDSSAYTTLLRTALFGPDVAGPYTPKTDSPSMTLPRNRIERYKRTETROSMSLSPMDDDVPGIINSPYKAPGRVRSPTKYLDAEPLQDDPYLINLVDSNRNVLAVLIGNKVIYIWMNCSKVKIKDIDGVDCVSGVAQRGHLDQDYLTINLVDSNRNVLAVLIGNKVIYIWMNCSKVKIKDIDGVDCVSGVAQRTIODEPVSFKLSGKHSLEVCGKWSYDNELASGGDNKLELVWNHSTQPVLYKCEHTAAVYKAIWSPHLHGLIASGGGTADRCIRFWMNTTSHLSCHMDTGSQVCNLWASKNVEIVSHGYSQNOIIVWRYPTMSKLATLTGHTYRVLYIAISPDQRTIVGAGDETLRFVNVPSPKQNTSEISIEGALISGRTTIR"					
BASE COUNT	2503 a	1387 c	1495 g	3065 t	
ORIGIN					
alignment_scores:					
Quality:	1779.00				Length: 1481
Ratio:	3.901				Gaps: 13
Percent Similarity:	30.790				Percent Identity: 30.520
alignment_block:					
US-09-701-572-2 x AF134835	..				
Align seg 1/1	to: AF134835	from: 1	to: 8450		
1	MetASPGLYThGIYASNAArgASnProProProthSerThrValaArgAS	17			
2461	ATTCGAGCGGAACCGGTATCGAATCCACACCGACTTCCACCGTCGAGGA	2510			
17	pasnSerProProProGluProSerProGluSerLeuAArgHISValSerA	34			
2511	TAATTTCACACCGCCGCGACCATCACCGGAAGTCTCCGCATCTAAAGCC	2560			
34	rgNellleasnSerAsnHISlyrThrSerProSerArgThIleTySer	50			
2561	GTAATGATCAACAGCAACCATTCACCTCATTTCGAAACATCTACTCC	2610			
51	AspArgPheIleProSerArgSerAlaSerIysPheAlaIeuPheAspI1	67			

```

|||||
2611 GATAGGTGATCCGAGTAGATCTGCTCGAATTCGCTTGTGGATAT 2660
67 easnThrProThrGluGlyArgAspSerSerSerAlaTyrThrL 84
|||||
2661 CAATACCTCGACGAGGAGGACGGCATGATAGTCCACCGCTTATACGAC 2710
84 euleuArgThrAlaLeuPheGlyProAspValAlaGlyProValThrPro 100
|||||
2711 TTCTTGAAACGGCGTGTGGACCGGATGTTCAGGTCCGGTTACGCCG 2760
101 GluLysThrAspSerProSerMetThrLeuProAsnArgAsnIlePheArg 117
|||||
2761 GAGAAAAACCGACTCGCCGTCGATGACATTCGCCAATAGCATATTTT 2810
117 gTyrLysThrGluThrArgGlnSerMetHisSerLeuSerProPheMet 134
|||||
2811 GATTAAGACGAGACGACAGACATCCATGCACTGCTTCGCCGTTATG 2860
134 spAspAspPheValProGlyValAsnHisSerProValLysAlaProArg 150
|||||
2861 ATGATGATTTGTCTCTGTATTAATCATAGTCCGTTAAGCTCTAGG 2910
151 LysValProArgSerProTyrLys 158
|||||
2911 AAGGTTCCCGATCGCCTTATAGGATTTCAATAAAAAAATTATCTT 2960
158 2961 TTTTATTTCTTGTGATCTTTGTTGAATGTTGATTTAGGCTTTTGA 3010
159 159 159 .....ValLeuAspAlaProAlaLeuGlnAspAspPheTyrLeu 171
|||||
3011 TTTTGATATAGCTTTGTGATGCACCTGCTTTCAGATGATTTTATCTG 3060
172 AsnLeuValAspTyrSerSerHisAsnValLeuAlaValGlyLeuGlyAs 188
|||||
3061 AATCGTAGATTTGCTTCACACAAATGTTGGCTGTTGGTAA 3110
188 nCysValTyrLeuTyrPasnAlaCysSerSerLys 199
|||||
3111 CGGTGTTATCTCTGGAATGCTTGTAGCAGCAAGGTGATTTTATCAT 3160
199 199 199 .....
3161 ATGATTATGTTGATTTCAAAATTTTCAATTTTATGATATGAG 3210
199 199 199 .....
3211 CTAGATGGTGTGGCTAGGAAGATTGAACCTGAAGTTGTTTTPA 3260
200 200 200 .....ValThrLysLeuCysA 205
|||||
3261 ATTGATTATTGATCTTGATATGTTTATGAGGTATTAATATATG 3310
205 spleuGlyValAspAspCysValCysSerValGlyTyrAlaGlnArgGly 221
|||||
3311 ATTTGGGGTGTGATGATGTTGTTGTTGTTGGCTCAACGCGCT 3360
222 ThrHisLeuAlaValGlyThrAsnAsnGlyLysValGln 234
|||||
3361 ACTCATCTTGGCTGTGGAACTAACAATGTAAAGTTCAAGTAAGATT 3410
234 234 234 .....
3411 TGTGCTAATGTTGATCATTAATAGGTTTCTAATGAGTGATTTGAT 3460
234 234 234 .....
3461 TGTATTGTGCATTAAGAGGTTCTTACGTTTATTTATTTGCTGGA 3510
234 234 234 .....

```

```

3511 TTTCTTCTGGGTTTCCCTGTGCTGTTTATTTTCTTCATATATG 3560
234 234 234 .....
3561 TTTCTTATTTTCCATTTGAAAGAAATCTTTTCCATTTGGGATGAC 3610
234 234 234 .....
3611 ACGCATTTGATGATTAACAGCATGTGTTATCTTACATTCGAATTA 3660
234 234 234 .....
3661 TTATAGCATTTTATCATTTCTAATATGTTGATTAACGTATTAAGT 3710
234 234 234 .....
3711 ATACTGAACCCATTAATTTTGGTAACAGCCCTTTTCTGTATGACAG 3760
234 234 234 .....
3761 GAGCTTCTGTTCTGGGTAGACTCTTAGACAGTATATACCTTCTAT 3810
234 234 234 .....
3811 ATGTCTAATACATATAGCATTAAGCATTAAGGATATATTATATATTA 3860
234 234 234 .....
3861 GAATTTATGTTTATGATTAATGCAATTAATGATGATACATTAATGTT 3910
234 234 234 .....
3911 TTGATAGTTTCTTGTATTAAGCTCTGTGATATTTGCCGTATGAGCT 3960
234 234 234 .....
3961 CTGTGCTGTTGCTGTGATTTGCTGATGATGCCGTGTGCTGTT 4010
234 234 234 .....
4011 GTGATTAATGCTATATTCATTCTAATGCTATGTTGATGCTATTA 4060
234 234 234 .....
4061 TTAATGCCGTGCTGCTGCTGTTGTGATTAATGCTATATTCATTCGTG 4110
234 234 234 .....
4111 CCGTGTGCTGATGCTATTAATTAATGCCGTGCTGTTGCAATTA 4160
234 234 234 .....
4161 GTCTATTAATTCCTGTGCTGTTGTCATTTGTCATTAATTAATGCCG 4210
234 234 234 .....
4211 TGTGCTGTGCTATATGCTATTAATTAATGCTGCTGCTGCTAT 4260
234 234 234 .....
4261 ATGCTATTAATTAATTTATTAATGCTAATCTTACTTCTTCT 4310
234 234 234 .....
4311 TGATCATCAAAATGTTGTTACTTAATAGTGGTCAATGTAACCGTAT 4360
234 234 234 .....
4361 ACCAGTAACACATCTGCATTAATTTATTAATTAATCAATGATTA 4410
234 234 234 .....
4411 GTTCAATTAATTTTGAATTCGAGCTTGTGCTTGTATAGCTTAC 4460

```

234 234
 4461 CATTACCCATGCGCAGTCTCTACATATGACCAGTGGCACTCA 4510
 234 234
 4511 TGTTCGATGTCAAACCATTTGGTTAAAAAATTAGTAGCAATACTT 4560
 234 234
 4561 CGATACATATGATGTGACTAGAACCAATCTTAACCAATATGATTA 4610
 234 234
 4611 GAGGGAGAGAGTCAAGTCTTTATGATATTCACCACTGAAACCACTG 4660
 234 234
 4661 ATGCATGATCATGCTCTGTAGCTAATCATTTAATGAATAGTTAAT 4710
 234 234
 4711 AATTAATGTTGTTCTGTAAGTCTTTTCTGGAAGAAAAAATGAGACG 4760
 234 234
 4761 GTTCTGTATTAATCTGTAATATCTGACGTCACACTGACCTGAGG 4810
 234 234
 4811 TTCTACAGATCTAGAAATGACGCTTTAATAAAGCGGCTTAAGCA 4860
 234 234
 4861 ACTTAGATCTGTTTGGCAACTGTAAGTCTGTAAGCTATCTACTG 4910
 234 234
 4911 AGATAGTCTATACCTTTAGGAGGCTCTATATTTGACTTTTAACGAGA 4960
 234 234
 4961 ACACCTAATTTGCTTGTGATGATCAATCTAATCAATCAACCATAC 5010
 234 234
 5011 ACATTCTCCCTCAAGAGTGTATATAGATTAACCTGTTTCTCTACCA 5060
 234 234
 5061 TAGTATATAGAAAGTGCATCACTTCATTAATTTTCCTTTGAGAC 5110
 234 234
 5111 TGCCTCCCATGAAAGTGGTATTTAGTGAATCTGCTTAACCTGTC 5160
 234 234
 5161 TGGTATCAAAATATGAATATCTATTTTCCATTTGCATCCCTGGAAT 5210
 234 234
 5211 TCGAAGTCTACTATTTCTTTTTTTAGGAAACTTAATGCTCTCTGTT 5260
 235 235
 5261 CTATGCTCAAAATTAATCAATGCTTTGTTTTTACTATATCAATGAT 5310
 235 235
 5311 TTGGGATGACGACAGATGCAAGATTAAGATCAATGAGGCGCATCGT 5360

252 euArgValGlyAlaLeuAlaIleuAlaIleuSerSerSerLeuSerSerGlyGly 268
 5361 TAGGTGTCGTCCTGCTGGCTGAGTTCATCTTTGTTGCTGCTGGA 5410
 269 ArgAspLysAsnIleTyrGlnArgAspIleArgThrGlnLysAspPheVa 285
 5411 CGGATAGAGATATTTATCAACGAGATATACGACACACAAAGATTTGT 5460
 285 IserLysLeuSerGlyHisLysSer..... 293
 5461 TAGTAACCTGTCAGACACAAATCAGAGGTAGATTAGTCTATTTCT 5510
 293 293
 5511 AGCTTCATCACTTATACATGCTATTCATCAATGATTTGATTTGTTT 5560
 294 Gln..ValCysGlyLeuLysTrpSerTyrAspAsnArgLeu 307
 5561 ATTTTTCGAAGTTTGTGAGCTGAAGTGTCAATATGATTAACGTGAGT 5610
 308 AlaSerGlyGlyAsnAspAsn..Lys..... 315
 5611 GCATCTGAGAAATGACAAACAAGTATGCAATCAACAACTTAATGATT 5660
 315 315
 5661 ATCTGAGATTCACACTGCATATTAATTTCTTTCTCTCTGACAGAA 5710
 316 LeuPheValIlePheAsnGlnHisSerTh 324
 5711 CTCACATCAATCATATTGTTGAGTGTGTTGTTGTTGATCAACACTCAAC 5760
 324 rGlnProValLeuLysTyrCysGlnHisThrAlaAlaValLysAlaIleA 341
 5761 CCAGCTGTCTCCTCAAGTACTGTGACACACAGCAGCTGTTAAAGTATTG 5810
 341 IatTrpSerProHisLeuHisGlyLeuAlaSerGlyGlyThrAla 357
 5811 CATGCTCTCTCATCTTCATGAGACTCTTCATCTGAGAGAGAACTGCA 5860
 358 AspArgCysIleArgPheTrpAsnThrThrAsnSerHisLeuSerCy 374
 5911 TATGACACTGGAAGTCAGTAAGTACTTCTTAACGTGTTATATTTT 5960
 380 380
 5961 CATACAAATGCACCTGATTTCAAAATTTGCTGTAATTTGGCGCGCTGT 6010
 380 380
 6011 TGTGGCCTTATATAGTTGTGTCTAGTGTCTAGAAACTATATTTAAT 6060
 380 380
 6061 TTGTTGGAAGGGAAGGGGAATGCTACTTCCCAACACCTTCCCTCT 6110
 380 380
 6111 CAGTTTTCCTCACTTACACAGCATTTTATGAGTGTATGAATATCAT 6160
 380 380
 6161 TCCTTACTTATCATATATAGTATGACTGACATCACTGGAAGAAATGCTC 6210
 380 380
 6211 CAGCTAGATTTTCTTACTATTAAGGTTGTAAGAAAAATCTTAATA 6260
 380 380

6261 AAACGAAAGCAGATCCCTAGTCTCTACTGTTACCATTCATATACATGCAAA 6310
380 380
6311 GTTAGATTATGGAAGTCTTTTTCAGTTCATTAAGTTTCAGATATATAGT 6360
380 380
6361 ATTTGCCCTTACATATTGAGAGAAGATTAATAGAGTCTTCTGCTGTTT 6410
381 Va 381
6411 CTCCTTTCTTAACACAAATCAATGATTTTATATCTATATATTTTGCAGT 6460
381 |CysAsnLeuValTTPSerLyAsnValAsnGluLeuValSerThrHisG 398
6461 TTGCATCTTGTCTGCTGCCAAAATGCAATGCAATGATAGTACACACATG 6510
398 |LYTysSerGlnAsnGlnIleIleValTTPArgTYRProThMetSerLys 414
6511 GCTACTCCACGACACCATTAATTTGTGGAGATACCCCATATGTCAAG 6560
414 414
6561 GTACAAAGTGTCTCTGATGAAGATGATGTCTTTGTTATACCACTAG 6610
415 LeuAlaThrL 418
6611 TTAACAAGTTTATCCATATTTTGTCTTGTGATGACAGCTGCGCAGTC 6660
418 |euthrGlnHisThrTYrArgValLeuTYrLeuAlaIleSerProAspGly 434
6661 TTACCGGCACACTTATAGGGTCTCTATCTGTGCATCTCCGAGATGA 6710
435 Gln..... 435
6711 |TTTCTTTTATTTAGTCTTCTTCACTAATCCTCATTTAGATTACTAGTATGT 6760
435 435
6761 AATGACAAAGAAATATTTTATATCATCTATGCTATATCTCGATTGTG 6810
436 Thr..IleValThrGlyAlaGlyAspGluThrLeuA 447
6811 ATGAGTACTATTTTCAGACTATTTGTAATGAGCTGAGATGAACGCTTA 6860
447 rghpethrAsnValIpheProSerProLysSerGln 458
6861 GGTTCTGGAATGTTTCCCTTCCCTTAATACACAG 6895
seq_name: gb_pl:ATF7H19
seq_documentation_block:
LOCUS ATF7H19 139316 bp DNA linear PLN 27-AUG-1999
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone F7H19 (ESSA
project).
ACCESSION AL031018
VERSION AL031018.1 GI:3292807
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 139316)
Beyan,M., Peters,S.A., van Staveren,M., Dirkse,W., Stiekema,W.,
Bancroft,I., Mewes,H.W., Mayer,K.F.X., Lemoine,K. and Schueller,C.
Unpublished
2 (bases 1 to 139316)
EU Arabidopsis sequencing project.
Direct Submission
Submitted (21-JUN-1999) MIPS, at the Max-Planck-Institut fuer

COMMENT

Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/ch14/Location/Qualifiers>

FEATURES

1000

source	location/Qualifiers
exon	1..139316 /organism="Arabidopsis thaliana" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4" complement(915..1445) /gene="F7H19.10" /number=1
CDS	complement(915..1445) /gene="F7H19.10" /note="similarity to PVPR3 protein, Phaseolus vulgaris, PID:g169363 contains EST gb:AA597458, H36610" /codon_start=1 /product="putative protein" /protein_id="CA19798.1" /db_xref="GI:3292808" /db_xref="SPTREMBL:O49663" /translation="MGSEQNDSTSTFQSOASEPKLCYKGCGFSPSNMDLCKCYNG ICAEEOPTAVAKAAVSEKSPSPSPSLFLAEPPAVVREPREKAAVYVNSPESSAV PEANESPRAPRNRCICCNKKYNGFCKGCTGCGHRIPEPHDCSFDKEVGE IAKANVYVADKIQR"
gene	complement(915..1445) /gene="F7H19.10"
gene	915..1445 /gene="F7H19.10"
exon	4680..6720 /gene="F7H19.20" complement(4680..4751) /gene="F7H19.20"
gene	/number=1 complement(join(4680..4751,4828..4910,5004..5257, 5355..5607,5685..5858,5956..6100,6319..6720)) /gene="F7H19.20"
CDS	complement(join(4680..4751,4828..4910,5004..5257, 5355..5607,5685..5858,5956..6100,6319..6720)) /gene="F7H19.20" /note="similarity to Na(+) dependent transporter, Aquifex aerolius, PIR2:E70482" /codon_start=1 /product="putative protein" /protein_id="CA19799.1" /db_xref="GI:3292809" /db_xref="SPTREMBL:O49665" /translation="MSVITPTIEHLIKSTLRLLPRAVVSQRIQVPPPISENTLS SPLRIDIPISGIFILSSITLRLCLITGCTKYTFEFGIQTPIITYGGRINLMRYA SDNSEKGLDEADPEKFCAMGSFLFHSMLCWVERPSIVDRKKKANSILPHVLA LLIALLYPSFWFTSREPLFQGLPMAVGNSENKDFLEFRPAKAILGLYGOVL KVLQIFIGLAVSLFQPLPTIGAGIMLVCSGSAQLSNATFRLTPALALSGIVTS LSTAAVLAIVPEMLSLILGKRLPDVKGMTSSIIQVYIAPILAGLILNKLPVNSAI RFLPLILVIDTACCVGALIKRLINNSVMSPEGATILLIYTMPLLSAFIAGYPLTGSV RNAAPRAKMQRTLSLETGMOSLLALALATKTFODPLVGIPIAISTVMSLKGFTLV IWSKESNTF"
Intron	complement(4752..4827) /gene="F7H19.20"
exon	/number=1 complement(4828..4910) /gene="F7H19.20"
Intron	/number=2 complement(4911..5003) /gene="F7H19.20"
exon	/number=2 complement(5004..5257) /gene="F7H19.20"

[illegible]

```

118 rLysThrGluThrArgGlnSerMetHisSerLeuSerProPheMetAspA 135
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
26161 TAAGACGGAGACTCATCGCTCTTGAATTCGTTCTCTCTTTGGCTTG 26210
135 spAspPheValProGlyValAsnHisSer...ProValLysAlaProArg 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
26211 ATGATGATTCCTCGTGGTTCTTCATAGTGCCTGTAAAGCTCCACG 26260
151 LysValProArgSerProTyrLysVal..... 159
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
26261 AAAGTCCCGGATCGCGTAAAGTGACGATTTTGATTCGGGGTGA 26310
159 ..... 159
26311 TACATGTTAGATTTTGGTGTTCATCATGAACTCGATCAGTCCTT 26360
159 ..... 159
26361 GTCGTTTAAAGTGGGTGTCGTCATGTCGAAATGATGTTTTTTC 26410
159 ..... 159
26411 TTCGATGGGTTTTTGTGTAATGCTGTAGATTGGGGGTGGAGCTTA 26460
159 ..... 159
26461 GATTCCTGATCTGCTGACTTATAGATCTTGGTTTCGATAATGATGAA 26510
159 ..... 159
26511 CAATTTGATCTTGTGATGATGTTTGGTCTGAGGCTCTAGAAATTTGAG 26560
159 ..... 159
26561 TCGAGATTAATGATCTTAGATGAACATATATTAGAAATGCTTGTTC 26610
159 ..... 159
26611 TTTTACTGTTCTAGTGTGTGTATATTAGAAAAGTTAAATCTGTT 26660
159 ..... 159
26661 ACTGGTTGGGTAAATTTGTTGGGCAATGTCCTACATCTGATPACA 26710
159 ..... 159
26711 GAGGTTTTCAGTACTTACTAGTGTAGAGTATCTAGATGATGACAA 26760
159 ..... 159
26761 TGTTCGTTGTTACGTTTTCGATAGTTTTCAGTGTGTGTTAAATG 26810
159 ..... 159
26811 CAAACCTGATGTTGTTATGATTTGGATTTGTTTATAGGCAATG 26860
160 ..... 160
26861 TTGCTACAGATGATTAACAGAGCTCTTCTGCTTTTATGATGATG 26910
162 IAProAlaLeuGlnAspAspPheTyrLeuAsnLeuValAspTyrSer 178
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
26911 CACCGGCTTCGACGATGATTTTATCTGAATCTTGGATTCGCTGCA 26960
179 HisAsnValLeuAlaValAlaGlyLeuGlnLysCysValTyrLeuTyrPasnAl 195
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
26961 CAAATGTTCTAGACAGTGGAGCTAGGAACGTGTATTTATGGAATGC 27010
195 acySerSerLys..... 199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27011 TTGTAGCAGCAAGTGAGTTCCTTTTTCACCTTGCAAAATTCATTAAT 27060

```

```

199 ..... 199
27061 CACAGTACTTCAAAATTTTGGTGGTTCGATTCCTCTTCAAAATC 27110
200 .....ValThrLysLeuCysAspLeuGlyValAspPheCysValCysSe 214
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27111 CAACAGGTACTAAGTTATGATCTCGAGCTCGAGATATGTTTCTC 27160
214 rValGlyTyrPalaglnArgGlyThrHisLeuAlaValAlaGlyThrAsnAsnG 231
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27161 AGTGGTTGGCGCTTACGTTGAGACTCATCTGCTGTGGAACTAGTACCG 27210
231 LysValGlnIle..... 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27211 GGAAGTTACAGTTAGATCATCTTACTCTTGTTCCTCTCTACTA 27260
235 ..... 235
27261 AGAATTTACGTTACTTTCGACACTTTCGAAATTTAGAAAACATATTTAAA 27310
235 ..... 235
27311 GTAACTATTTATAGTACAAACATGATAGATTAGACCTTATCGTGCAC 27360
235 ..... 235
27361 TGTAGGATTCCTGGCAATATCTTTACTGCTATGATCTTGATCATTAAGTCG 27410
235 ..... 235
27411 TATGATTTATCAAAAGATTTTCTTATGCGAGAGATATCTCGATTA 27460
235 ..... 235
27461 AGCATTTCAAGGCTTGTAAATTTTGAAAGTCCCTGAATGAACTA 27510
235 ..... 235
27511 AATCCTTTGTGCTTAAGATGGAAGTATCTGCATATAAACCTTGGCGGA 27560
235 ..... 235
27561 TAAATTTTCATATATCTGAAGAAATTTATGATGTTCTTCTATAGCATAT 27610
236 ..... 236
27611 TCACCTTAGTTCGTTTTCGATATGAGATGCGTCACGCTCAGAGAGA 27660
244 IleArgSerMetGlnGlyHisArgLeuArgValAlaLeuAlaTyrPse 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27661 ACAAGAACATGGAAGGTCAATCGTCTAAGATTGGAAGCCCTGACATGGCG 27710
260 rSerSerLeuLeuSerSerGlyArgAspLysAsnIleTyrGlnArgA 277
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27711 TTCATCGTTCCTGCATCTGTAGCAGACAGCAAGAGATTCCTCAGAGAG 27760
277 spIleArgThrGlnGlnAspPheValSerLysSerGlyHisLysSer 293
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27761 ACATTAAGCTGCAAGAGATCATGTCAGTAAATTTGGAGCTCAATTAATCT 27810
294 GluValCysGlyLeuLysTyrSerTyrAspAsnArgGlnLeuAlaSerG 310
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27811 GAGGTATGCGGACCTCAAGTGTCTTATGACAAACAGAGAGCTACATCTGG 27860
310 yGlyAsnAspAsnLys..... 315
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27861 TGGAAACGACAAATAGGTATGATGATCATGATTTGGGTTACATTAG 27910
315 ..... 315
27911 TCTTCTTTGTAACCCCAACATAGCTAAATGCTTATTAATTTTCAG 27960
316 LeuPheValTyrPasnGlnHisSerThrGlnProValLeuLysTyrCysG 332

```

```

|||||
27961 CTTTGTGATGGACACATTCAACACACCGGTTTGAATATAGTCA 28010
332 UHSTHRAAAlaValLysAlaIleAlaIlePseProHISLeuHISGLYL 349
|||||
28011 ACACACTGACGCTTAAAGCCATGCTGCTCCTCATCTTCAATGGCG 28060
349 euleuAlaSerGlyGlyThrAlaAspArgCysIleArgPheTrpAsn 365
|||||
28061 TTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 28110
366 ThrThrAsnSerHisLeuSerCysMetAspThrGlySer..... 379
|||||
28111 ACACACGCAATATCAATTAAGTTCATAGTACTGCAGTCAGGTAAA 28160
379 ..... 379
28161 TTTCTTAATGATGATGATGATGATGATGATGATGATGATGATGATG 28210
380 ..... GluValCysAsnLeuValTrp 386
28211 TTAATATCATCTTTGCTTCTATATATTAACAGTATGCAATCTAGCT 28260
387 SerLysAsnValAsnGluLeuValSerThrHisGlyTrpSerGlnAsnG 403
|||||
28261 TCTAAGACGTAACGAGCTTGTAGACACACGAGATCTCTCAGAACCA 28310
403 nIleIleValTrpArgTrpProThrMetSerLys..... 414
|||||
28311 AATCATATGCTGGAATATACCAACATGTCGCAATGAAATCTCTCGA 28360
414 ..... 414
28361 CCTATATATCTTCCATCATGATGATGATGATGATGATGATGATGAT 28410
415 ..... LeuAlaThrLeuThrGlyHisThrTrpArgValIle 426
|||||
28411 TTTTGGTTGTTGATGATGATGATGATGATGATGATGATGATGATG 28460
426 uTrpLeuAlaIleSerProAspGlyGlnThrIleValThrGlyAlaG 443
|||||
28461 ATACCTGGGTTTCACCGCATGACAGACGATGTAACAGAGAGAGAG 28510
443 spGluThrLeuArgPheTrpAsnValPheProSerProLysSer.... 457
|||||
28511 ATGAAACCTTAAGGTTCTGGAATGTTTCCCTCCCAAAATCTCAGGTA 28560
457 ..... 457
28561 AAAACCACTACTCTGTTGTAATTACTTAATGATCACAATAATGCT 28610
458 ..... GluAsnThrG 461
28611 ATTTCTCAGAAAGCTGATTCGATGTTTGGCTGTTGTTGCAACACG 28660
461 luserGluIleGlyAlaLeuSerLeuGlyArgThrThrIleArg 475
|||||
28661 ATAGTGAATCGGTCGCTTTCTTGGTAGAACAACAAATTGCG 28704

```

seq_name: gb_pl:ATCHRIV58

seq_documentation block:

```

LOCUS ATCHRIV58 193550 bp DNA linear PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 58.
ACCESSION AL161558
VERSION AL161558.2 GI:7269123
KEYWORDS
SOURCE
thalid cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 29088 to 168403)

```

AUTHORS Peters, S.A., van Staveren, M., Dirkse, W., Stiekema, W., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.

REFERENCE JOURNAL Unpublished

AUTHORS 2 (bases 154397 to 193550)

REFERENCE JOURNAL Unpublished

AUTHORS 3 (bases 1 to 38442)

REFERENCE JOURNAL Unpublished

AUTHORS Hilbert, H., Braun, M., Holzer, E., Brandt, A., Duesterhoeft, A., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.

REFERENCE JOURNAL Unpublished

AUTHORS 4 (bases 1 to 193550)

REFERENCE JOURNAL EU Arabidopsis sequencing, project.

AUTHORS Direct Submission

TITLE Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de, project coordinator: Mike Beyer, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UD Norwich, UK, E-mail: michael.bevan@bsrc.ac.uk

COMMENT Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/> this fragment has an overlap with ATCHRIV57 at the 5' end and an overlap with ATCHRIV59 at the 3' end.

FEATURES

source Location/Qualifiers

1..193550

/organism="Arabidopsis thaliana"

/variety="Columbia"

/db_xref="taxon:3702"

/chromosome="4"

6009..7569

/gene="AT4g22770"

complement(6009..6398)

/gene="AT4g22770"

number=1

complement(join(6009..6398,6575..6706,6791..6858,7155..7569))

/gene="AT4g22770"

complement(join(6009..6398,6575..6706,6791..6858,7155..7569))

/note="similarity to DNA-binding protein PDL, Pisum sativum, PARCH:E275188 contains EST gb:N37194"

/codon_start=1

/product="putative DNA binding protein"

/protein_id="CAW79232.1"

/db_xref="GI:7269124"

/translation="METHGEVYKTTTGSDDGYVYRSNAPDFHMAPRSETSNTPNS VAPPPPPONSFTPSAAMDGFSSGPIKRRGRPKYGHDAAVTLSPNPISAPPT SHVIDFTTSSEKRMKRPATPPSSFFIRPKYOVETLGEWSPSSAANPTPIITVNA GEDVKRILISFOGSLAICLVANGVSVTLRQDSSGGTLTEGREILISGTFM PSDSGTFRSGPLMPTSSNVADHRTIRPMTSLPTSTWTFPPSDSRKHSHDNITTL KPHNNFMPSPLMPTSSNVADHRTIRPMTSLPTSTWTFPPSDSRKHSHDNITTL"

complement(6399..6574)

/gene="AT4g22770"

number=1

complement(6575..6706)

/gene="AT4g22770"

number=2

complement(6707..6790)

/gene="AT4g22770"

number=2

complement(6791..6858)

/gene="AT4g22770"

number=3

complement(6859..7154)

/gene="AT4g22770"

number=3

complement(7155..7569)

/gene="AT4g22770"

exon

intron

```

exon      /number=4
           complement(10826..11212)
           /gene="At4g22780"
           /number=1
gene       /gene="At4g22780"
           complement(join(10826..11212,11306..11977,12062..12129,
           12220..12334,12424..12496,13025..13086))
           /gene="At4g22780"
           complement(join(10826..11212,11306..11977,12062..12129,
           12220..12334,12424..12496,13025..13086))
           /gene="At4g22780"
           /note="similarity to translation factor EF-1 alpha
           gentfamily, Arabidopsis thaliana, PATCHX:G1532164
           contains EST gb:217780"
           /codon_start=1
           /product="translation factor EF-1 alpha-like protein"
           /protein_id="CAB79233.1"
           /db_xref="GI:7269125"
           /translation="MELSDCSNEYKLVYRNMPRVVIDNGVCPNSTVYVYFETLTPS
           SIDASRPGILLESVOLFTDMNLTKKAVISSDGRKMDVFNVDLNGDKLIDENLIR
           YLEKSTESHVCKTGGYNGITALELGTGDRVGLSEVFAVLADIEDCVYEAKEAMTHNG
           RIASMTYKDGNSGPIIDSDSPRVQVRBQGLNKLKADGYNDRPTCYSGKTHME
           RLHQRMEKDRDYERKEKESPIVSVQNLPRGYSVNLQCKDMKLEFDVYCTLD
           MAYIVFHAIRTVGEAFLEFVYRHSDGHPVSEPERQLDQDAIERTYKGVRL
           ELCTADRPGLAEVTRILRENGLNIAKEISTKDSIARNVFVTDANGNLIDPEIKS
           IREKIGIDLSVKEPPEPISCREAVEKEQHIEPDHQGRYGGTVLVSGLVMRLYH
           LGLIKSYF"
           complement(11213..11305)
           /gene="At4g22780"
           /number=1
exon      /complement(11306..11977)
           /gene="At4g22780"
           /number=2
           complement(11978..12061)
           /gene="At4g22780"
           /number=2
           complement(12062..12129)
           /gene="At4g22780"
           /number=3
           complement(12130..12219)
           /gene="At4g22780"
           /number=3
           complement(12220..12334)
           /gene="At4g22780"
           /number=4
           complement(12335..12423)
           /gene="At4g22780"
           /number=4
           complement(12424..12496)
           /gene="At4g22780"
           /number=5
           complement(12497..13024)
           /gene="At4g22780"
           /number=5
           complement(13025..13086)
           /gene="At4g22780"
           /number=6
           complement(13087..13175)
           /gene="At4g22780"
           /number=1
           complement(13176..13264)
           /gene="At4g22780"
           /number=1
           complement(13265..13353)
           /gene="At4g22780"
           /number=1
           complement(13354..13442)
           /gene="At4g22780"
           /number=1
           complement(13443..13531)
           /gene="At4g22780"
           /number=1
           complement(13532..13620)
           /gene="At4g22780"
           /number=1
           complement(13621..13709)
           /gene="At4g22780"
           /number=1
           complement(13710..13798)
           /gene="At4g22780"
           /number=1
           complement(13799..13887)
           /gene="At4g22780"
           /number=1
           complement(13888..13976)
           /gene="At4g22780"
           /number=1
           complement(13977..14065)
           /gene="At4g22780"
           /number=1
           complement(14066..14154)
           /gene="At4g22780"
           /number=1
           complement(14155..14243)
           /gene="At4g22780"
           /number=1
           complement(14244..14332)
           /gene="At4g22780"
           /number=1
           complement(14333..14421)
           /gene="At4g22780"
           /number=1
           complement(14422..14510)
           /gene="At4g22780"
           /number=1
           complement(14511..14599)
           /gene="At4g22780"
           /number=1
           complement(14600..14688)
           /gene="At4g22780"
           /number=1
           complement(14689..14777)
           /gene="At4g22780"
           /number=1
           complement(14778..14866)
           /gene="At4g22780"
           /number=1
           complement(14867..14955)
           /gene="At4g22780"
           /number=1
           complement(14956..15044)
           /gene="At4g22780"
           /number=1
           complement(15045..15133)
           /gene="At4g22780"
           /number=1
           complement(15134..15222)
           /gene="At4g22780"
           /number=1
           complement(15223..15311)
           /gene="At4g22780"
           /number=1
           complement(15312..15400)
           /gene="At4g22780"
           /number=1
           complement(15401..15489)
           /gene="At4g22780"
           /number=1
           complement(15490..15578)
           /gene="At4g22780"
           /number=1
           complement(15579..15667)
           /gene="At4g22780"
           /number=1
           complement(15668..15756)
           /gene="At4g22780"
           /number=1
           complement(15757..15845)
           /gene="At4g22780"
           /number=1
           complement(15846..15934)
           /gene="At4g22780"
           /number=1
           complement(15935..16023)
           /gene="At4g22780"
           /number=1
           complement(16024..16112)
           /gene="At4g22780"
           /number=1
           complement(16113..16201)
           /gene="At4g22780"
           /number=1
           complement(16202..16290)
           /gene="At4g22780"
           /number=1
           complement(16291..16379)
           /gene="At4g22780"
           /number=1
           complement(16380..16468)
           /gene="At4g22780"
           /number=1
           complement(16469..16557)
           /gene="At4g22780"
           /number=1
           complement(16558..16646)
           /gene="At4g22780"
           /number=1
           complement(16647..16735)
           /gene="At4g22780"
           /number=1
           complement(16736..16824)
           /gene="At4g22780"
           /number=1
           complement(16825..16913)
           /gene="At4g22780"
           /number=1
           complement(16914..17002)
           /gene="At4g22780"
           /number=1
           complement(17003..17091)
           /gene="At4g22780"
           /number=1
           complement(17092..17180)
           /gene="At4g22780"
           /number=1
           complement(17181..17269)
           /gene="At4g22780"
           /number=1
           complement(17270..17358)
           /gene="At4g22780"
           /number=1
           complement(17359..17447)
           /gene="At4g22780"
           /number=1
           complement(17448..17536)
           /gene="At4g22780"
           /number=1
           complement(17537..17625)
           /gene="At4g22780"
           /number=1
           complement(17626..17714)
           /gene="At4g22780"
           /number=1
           complement(17715..17803)
           /gene="At4g22780"
           /number=1
           complement(17804..17892)
           /gene="At4g22780"
           /number=1
           complement(17893..17981)
           /gene="At4g22780"
           /number=1
           complement(17982..18070)
           /gene="At4g22780"
           /number=1
           complement(18071..18159)
           /gene="At4g22780"
           /number=1
           complement(18160..18248)
           /gene="At4g22780"
           /number=1
           complement(18249..18337)
           /gene="At4g22780"
           /number=1
           complement(18338..18426)
           /gene="At4g22780"
           /number=1
           complement(18427..18515)
           /gene="At4g22780"
           /number=1
           complement(18516..18604)
           /gene="At4g22780"
           /number=1
           complement(18605..18693)
           /gene="At4g22780"
           /number=1
           complement(18694..18782)
           /gene="At4g22780"
           /number=1
           complement(18783..18871)
           /gene="At4g22780"
           /number=1
           complement(18872..18960)
           /gene="At4g22780"
           /number=1
           complement(18961..19049)
           /gene="At4g22780"
           /number=1
           complement(19050..19138)
           /gene="At4g22780"
           /number=1
           complement(19139..19227)
           /gene="At4g22780"
           /number=1
           complement(19228..19316)
           /gene="At4g22780"
           /number=1
           complement(19317..19405)
           /gene="At4g22780"
           /number=1
           complement(19406..19494)
           /gene="At4g22780"
           /number=1
           complement(19495..19583)
           /gene="At4g22780"
           /number=1
           complement(19584..19672)
           /gene="At4g22780"
           /number=1
           complement(19673..19761)
           /gene="At4g22780"
           /number=1
           complement(19762..19850)
           /gene="At4g22780"
           /number=1
           complement(19851..19939)
           /gene="At4g22780"
           /number=1
           complement(19940..20028)
           /gene="At4g22780"
           /number=1
           complement(20029..20117)
           /gene="At4g22780"
           /number=1
           complement(20118..20206)
           /gene="At4g22780"
           /number=1
           complement(20207..20295)
           /gene="At4g22780"
           /number=1
           complement(20296..20384)
           /gene="At4g22780"
           /number=1
           complement(20385..20473)
           /gene="At4g22780"
           /number=1
           complement(20474..20562)
           /gene="At4g22780"
           /number=1
           complement(20563..20651)
           /gene="At4g22780"
           /number=1
           complement(20652..20740)
           /gene="At4g22780"
           /number=1
           complement(20741..20829)
           /gene="At4g22780"
           /number=1
           complement(20830..20918)
           /gene="At4g22780"
           /number=1
           complement(20919..21007)
           /gene="At4g22780"
           /number=1
           complement(21008..21096)
           /gene="At4g22780"
           /number=1
           complement(21097..21185)
           /gene="At4g22780"
           /number=1
           complement(21186..21274)
           /gene="At4g22780"
           /number=1
           complement(21275..21363)
           /gene="At4g22780"
           /number=1
           complement(21364..21452)
           /gene="At4g22780"
           /number=1
           complement(21453..21541)
           /gene="At4g22780"
           /number=1
           complement(21542..21630)
           /gene="At4g22780"
           /number=1
           complement(21631..21719)
           /gene="At4g22780"
           /number=1
           complement(21720..21808)
           /gene="At4g22780"
           /number=1
           complement(21809..21897)
           /gene="At4g22780"
           /number=1
           complement(21898..21986)
           /gene="At4g22780"
           /number=1
           complement(21987..22075)
           /gene="At4g22780"
           /number=1
           complement(22076..22164)
           /gene="At4g22780"
           /number=1
           complement(22165..22253)
           /gene="At4g22780"
           /number=1
           complement(22254..22342)
           /gene="At4g22780"
           /number=1
           complement(22343..22431)
           /gene="At4g22780"
           /number=1
           complement(22432..22520)
           /gene="At4g22780"
           /number=1
           complement(22521..22609)
           /gene="At4g22780"
           /number=1
           complement(22610..22698)
           /gene="At4g22780"
           /number=1
           complement(22699..22787)
           /gene="At4g22780"
           /number=1
           complement(22788..22876)
           /gene="At4g22780"
           /number=1
           complement(22877..22965)
           /gene="At4g22780"
           /number=1
           complement(22966..23054)
           /gene="At4g22780"
           /number=1
           complement(23055..23143)
           /gene="At4g22780"
           /number=1
           complement(23144..23232)
           /gene="At4g22780"
           /number=1
           complement(23233..23321)
           /gene="At4g22780"
           /number=1
           complement(23322..23410)
           /gene="At4g22780"
           /number=1
           complement(23411..23499)
           /gene="At4g22780"
           /number=1
           complement(23500..23588)
           /gene="At4g22780"
           /number=1
           complement(23589..23677)
           /gene="At4g22780"
           /number=1
           complement(23678..23766)
           /gene="At4g22780"
           /number=1
           complement(23767..23855)
           /gene="At4g22780"
           /number=1
           complement(23856..23944)
           /gene="At4g22780"
           /number=1
           complement(23945..24033)
           /gene="At4g22780"
           /number=1
           complement(24034..24122)
           /gene="At4g22780"
           /number=1
           complement(24123..24211)
           /gene="At4g22780"
           /number=1
           complement(24212..24300)
           /gene="At4g22780"
           /number=1
           complement(24301..24389)
           /gene="At4g22780"
           /number=1
           complement(24390..24478)
           /gene="At4g22780"
           /number=1
           complement(24479..24567)
           /gene="At4g22780"
           /number=1
           complement(24568..24656)
           /gene="At4g22780"
           /number=1
           complement(24657..24745)
           /gene="At4g22780"
           /number=1
           complement(24746..24834)
           /gene="At4g22780"
           /number=1
           complement(24835..24923)
           /gene="At4g22780"
           /number=1
           complement(24924..25012)
           /gene="At4g22780"
           /number=1
           complement(25013..25101)
           /gene="At4g22780"
           /number=1
           complement(25102..25190)
           /gene="At4g22780"
           /number=1
           complement(25191..25279)
           /gene="At4g22780"
           /number=1
           complement(25280..25368)
           /gene="At4g22780"
           /number=1
           complement(25369..25457)
           /gene="At4g22780"
           /number=1
           complement(25458..25546)
           /gene="At4g22780"
           /number=1
           complement(25547..25635)
           /gene="At4g22780"
           /number=1
           complement(25636..25724)
           /gene="At4g22780"
           /number=1
           complement(25725..25813)
           /gene="At4g22780"
           /number=1
           complement(25814..25902)
           /gene="At4g22780"
           /number=1
           complement(25903..25991)
           /gene="At4g22780"
           /number=1
           complement(25992..26080)
           /gene="At4g22780"
           /number=1
           complement(26081..26169)
           /gene="At4g22780"
           /number=1
           complement(26170..26258)
           /gene="At4g22780"
           /number=1
           complement(26259..26347)
           /gene="At4g22780"
           /number=1
           complement(26348..26436)
           /gene="At4g22780"
           /number=1
           complement(26437..26525)
           /gene="At4g22780"
           /number=1
           complement(26526..26614)
           /gene="At4g22780"
           /number=1
           complement(26615..26703)
           /gene="At4g22780"
           /number=1
           complement(26704..26792)
           /gene="At4g22780"
           /number=1
           complement(26793..26881)
           /gene="At4g22780"
           /number=1
           complement(26882..26970)
           /gene="At4g22780"
           /number=1
           complement(26971..27059)
           /gene="At4g22780"
           /number=1
           complement(27060..27148)
           /gene="At4g22780"
           /number=1
           complement(27149..27237)
           /gene="At4g22780"
           /number=1
           complement(27238..27326)
           /gene="At4g22780"
           /number=1
           complement(27327..27415)
           /gene="At4g22780"
           /number=1
           complement(27416..27504)
           /gene="At4g22780"
           /number=1
           complement(27505..27593)
           /gene="At4g22780"
           /number=1
           complement(27594..27682)
           /gene="At4g22780"
           /number=1
           complement(27683..27771)
           /gene="At4g22780"
           /number=1
           complement(27772..27860)
           /gene="At4g22780"
           /number=1
           complement(27861..27949)
           /gene="At4g22780"
           /number=1
           complement(27950..28038)
           /gene="At4g22780"
           /number=1
           complement(28039..28127)
           /gene="At4g22780"
           /number=1
           complement(28128..28216)
           /gene="At4g22780"
           /number=1
           complement(28217..28305)
           /gene="At4g22780"
           /number=1
           complement(28306..28394)
           /gene="At4g22780"
           /number=1
           complement(28395..28483)
           /gene="At4g22780"
           /number=1
           complement(28484..28572)
           /gene="At4g22780"
           /number=1
           complement(28573..28661)
           /gene="At4g22780"
           /number=1
           complement(28662..28750)
           /gene="At4g22780"
           /number=1
           complement(28751..28839)
           /gene="At4g22780"
           /number=1
           complement(28840..28928)
           /gene="At4g22780"
           /number=1
           complement(28929..29017)
           /gene="At4g22780"
           /number=1
           complement(29018..29106)
           /gene="At4g22780"
           /number=1
           complement(29107..29195)
           /gene="At4g22780"
           /number=1
           complement(29196..29284)
           /gene="At4g22780"
           /number=1
           complement(29285..29373)
           /gene="At4g22780"
           /number=1
           complement(29374..29462)
           /gene="At4g22780"
           /number=1
           complement(29463..29551)
           /gene="At4g22780"
           /number=1
           complement(29552..29640)
           /gene="At4g22780"
           /number=1
           complement(29641..29729)
           /gene="At4g22780"
           /number=1
           complement(29730..29818)
           /gene="At4g22780"
           /number=1
           complement(29819..29907)
           /gene="At4g22780"
           /number=1
           complement(29908..29996)
           /gene="At4g22780"
           /number=1
           complement(30000..30088)
           /gene="At4g22780"
           /number=1
           complement(30089..30177)
           /gene="At4g22780"
           /number=1
           complement(30178..30266)
           /gene="At4g22780"
           /number=1
           complement(30267..30355)
           /gene="At4g22780"
           /number=1
           complement(30356..30444)
           /gene="At4g22780"
           /number=1
           complement(30445..30533)
           /gene="At4g22780"
           /number=1
           complement(30534..30622)
           /gene="At4g22780"
           /number=1
           complement(30623..30711)
           /gene="At4g22780"
           /number=1
           complement(30712..30800)
           /gene="At4g22780"
           /number=1
           complement(30801..30889)
           /gene="At4g22780"
           /number=1
           complement(30890..30978)
           /gene="At4g22780"
           /number=1
           complement(30979..31067)
           /gene="At4g22780"
           /number=1
           complement(31068..31156)
           /gene="At4g22780"
           /number=1
           complement(31157..31245)
           /gene="At4g22780"
           /number=1
           complement(31246..31334)
           /gene="At4g22780"
           /number=1
           complement(31335..31423)
           /gene="At4g22780"
           /number=1
           complement(31424..31512)
           /gene="At4g22780"
           /number=1
           complement(31513..31601)
           /gene="At4g22780"
           /number=1
           complement(31602..31690)
           /gene="At4g22780"
           /number=1
           complement(31691..31779)
           /gene="At4g22780"
           /number=1
           complement(31780..31868)
           /gene="At4g22780"
           /number=1
           complement(31869..31957)
           /gene="At4g22780"
           /number=1
           complement(31958..32046)
           /gene="At4g22780"
           /number=1
           complement(32047..32135)
           /gene="At4g22780"
           /number=1
           complement(32136..32224)
           /gene="At4g22780"
           /number=1
           complement(32225..32313)
           /gene="At4g22780"
           /number=1
           complement(32314..32402)
           /gene="At4g22780"
           /number=1
           complement(32403..32491)
           /gene="At4g22780"
           /number=1
           complement(32492..32580)
           /gene="At4g22780"
           /number=1
           complement(32581..32669)
           /gene="At4g22780"
           /number=1
           complement(32670..32758)
           /gene="At4g22780"
           /number=1
           complement(32759..32847)
           /gene="At4g22780"
           /number=1
           complement(32848..32936)
           /gene="At4g22780"
           /number=1
           complement(32937..33025)
           /gene="At4g22780"
           /number=1
           complement(33026..33114)
           /gene="At4g22780"
           /number=1
           complement(33115..33203)
           /gene="At4g22780"
           /number=1
           complement(33204..33292)
           /gene="At4g22780"
           /number=1
           complement(33293..33381)
           /gene="At4g22780"
           /number=1
           complement(33382..33470)
           /gene="At4g22780"
           /number=1
           complement(33471..33559)
           /gene="At4g22780"
           /number=1
           complement(33560..33648)
           /gene="At4g22780"
           /number=1
           complement(33649..33737)
           /gene="At4g22780"
           /number=1
           complement(33738..33826)
           /gene="At4g22780"
           /number=1
           complement(33827..33915)
           /gene="At4g22780"
           /number=1
           complement(33916..34004)
           /gene="At4g22780"
           /number=1
           complement(34005..34093)
           /gene="At4g22780"
           /number=1
           complement(34094..34182)
           /gene="At4g22780"
           /number=1
           complement(34183..34271)
           /gene="At4g22780"
           /number=1
           complement(34272..34360)
           /gene="At4g22780"
           /number=1
           complement(34361..34449)
           /gene="At4g22780"
           /number=1
           complement(34450..34538)
           /gene="At4g22780"
           /number=1
           complement(34539..34627)
           /gene="At4g22780"
           /number=1
           complement(34628..34716)
           /gene="At4g22780"
           /number=1
           complement(34717..34805)
           /gene="At4g22780"
           /number=1
           complement(34806..34894)
           /gene="At4g22780"
           /number=1
           complement(34895..34983)
           /gene="At4g22780"
           /number=1
           complement(34984..35072)
           /gene="At4g22780"
           /number=1
           complement(35073..35161)
           /gene="At4g22780"
           /number=1
           complement(35162..35250)
           /gene="At4g22780"
           /number=1
           complement(35251..35339)
           /gene="At4g22780"
           /number=1
           complement(35340..35428)
           /gene="At4g22780"
           /number=1
           complement(35429..35517)
           /gene="At4g22780"
           /number=1
           complement(35518..35606)
           /gene="At4g22780"
           /number=1
           complement(35607..35695)
           /gene="At4g22780"
           /number=1
           complement(35696..35784)
           /gene="At4g22780"
           /number=1
           complement(35785..35873)
           /gene="At4g22780"
           /number=1
           complement(35874..35962)
           /gene="At4g22780"
           /number=1
           complement(35963..36051)
           /gene="At4g22780"
           /number=1
           complement(36052..36140)
           /gene="At4g22780"
           /number=1
           complement(36141..36229)
           /gene="At4g22780"
           /number=1
           complement(36230..36318)
           /gene="At4g22780"
           /number=1
           complement(36319..36407)
           /gene="At4g22780"
           /number=1
           complement(36408..36496)
           /gene="At4g22780"
           /number=1
           complement(36497..36585)
           /gene="At4g22780"
           /number=1
           complement(36586..36674)
           /gene="At4g22780"
           /number=1
           complement(36675..36763)
           /gene="At4g22780"
           /number=1
           complement(36764..36852)
           /gene="At4g22780"
           /number=1
           complement(36853..36941)
           /gene="At4g22780"
           /number=1
           complement(36942..37030)
           /gene="At4g22780"
           /number=1
           complement(37031..37119)
           /gene="At4g22780"
           /number=1
           complement(37120..37208)
           /gene="At4g22780"
           /number=1
           complement(37209..37297)
           /gene="At4g22780"
           /number=1
           complement(37298..37386)
           /gene="At4g22780"
           /number=1
           complement(37387..37475)
           /gene="At4g22780"
           /number=1
           complement(37476..37564)
           /gene="At4g22780"
           /number=1
           complement(37565..37653)
           /gene="At4g22780"
           /number=1
           complement(37654..37742)
           /gene="At4g22780"
           /number=1
           complement(37743..37831)
           /gene="At4g22780"
           /number=1
           complement(37832..37920)
           /gene="At4g22780"
           /number=1
           complement(37921..38009)
           /gene="At4g22780"
           /number=1
           complement(38010..38098)
           /gene="At4g22780"
           /number=1
           complement(38099..38187)
           /gene="At4g22780"
           /number=1
           complement(38188..38276)
           /gene="At4g22780"
           /number=1
           complement(38277..38365)
           /gene="At4g22780"
           /number=1
           complement(38366..38454)
           /gene="At4g22780"
           /number=1
           complement(38455..38543)
           /gene="At4g22780"
           /number=1
           complement(38544..38632)
           /gene="At4g22780"
           /number=1
           complement(38633..38721)
           /gene="At4g22780"
           /number=1
           complement(38722..38810)
           /gene="At4g22780"
           /number=1
           complement(38811..38899)
           /gene="At4g22780"
           /number=1
           complement(38900..38988)
           /gene="At4g22780"
           /number=1
           complement(38989..39077)
           /gene="At4g22780"
           /number=1
           complement(39078..39166)
           /gene="At4g22780"
           /number=1
           complement(39167..39255)
           /gene="At4g22780"
           /number=1
           complement(39256..39344)
           /gene="At4g22780"
           /number=1
           complement(39345..39433)
           /gene="At4g22780"
           /number=1
           complement(39434..39522)
           /gene="At4g22780"
           /number=1
           complement(39523..39611)
           /gene="At4g22780"
           /number=1
           complement(39612..39700)
           /gene="At4g22780"
           /number=1
           complement(39701..39789)
           /gene="At4g22780"
           /number=1
           complement(39790..39878)
           /gene="At4g22780"
           /number=1
           complement(39879..39967)
           /gene="At4g22780"
           /number=1
           complement(39968..40056)
           /gene="At4g22780"
           /number=1
           complement(40057..40145)
           /gene="At4g22780"
           /number=1
           complement(40146..40234)
           /gene="At4g22780"
           /number=1
           complement(40235..40323)
           /gene="At4g22780"
           /number=1
           complement(40324..40412)
           /gene="At4g22780"
           /number=1
           complement(40413..40501)
           /gene="At4g22780"
           /number=1
           complement(40502..40590)
           /gene="At4g22780"
           /number=1
           complement(40591..40679)
           /gene="At4g22780"
           /number=1
           complement(40680..40768)
           /gene="At4g22780"
           /number=1
           complement(40769..40857)
           /gene="At4g22780"
           /number=1
           complement(40858..4
```

85 euargThrAlaLeuPheGlyProAspValAlaGlyProValThrProGlu 101
 |||||.....
 55166 TGGGTGGGGAGTGGTCTGAG.....ACGCCGGAG 55200
 102 LysThrAspSerProSerMetThrLeuProAsnArgAsnIlePheArgTy 118
 |||||.....
 55201 AAGAGAGATATTACTGGGTTTCT...TCTCCAGAGATATTATTAGGTT 55247
 118 rLysThrGluThrArgGlnSerMetHisSerLeuSerProPheMetAspA 135
 :|||.....
 55248 TAAGACGAGAACATCATCGCTTGAATTCGTTTCTCCTTTGGGTGG 55297
 135 spAspPheValProGlyValAsnHisSer...ProValLysAlaProArg 150
 |||||.....
 55298 ATGATGATTTCTCCGGTGTTCATAGTAGTCTCTGTTAAAGCTCCAGG 55347
 151 LysValProArgSerProTyLysVal..... 159
 55348 AAAGTCCGGGATCGCGGTAAAGTGAGCATTTTGATTCGGGGTGA 55397
 159 159
 55398 TACATTTGATTGTTTGGTGTGTTGATCATGTGAATCGATCATGCTT 55447
 159 159
 55448 GTGTTTTTAAGTGGTGTGTGCATGTGTGGAATTGATTGTTTTTTC 55497
 159 159
 55498 TTCGATGGGTTTTGTTGTAATGCTGTAGATTGGGGGTTGGGAGCTTA 55547
 159 159
 55548 GATTTCTGATCTGTTGACTTTAGATCTTGTTGTTGATATGCAAGAA 55597
 159 159
 55598 CAATTTGATCTTTGATGATTTTGGTCTCTGAGGCTAGAAATTGAG 55647
 159 159
 55648 TCTGAGGTTAATGATCTTAGATGAACATATTATGAAGATGCTTGTG 55697
 159 159
 55698 TTTTACTGTTCTTAGTGTGTGTATATTAGAAAGTTAAATCTTGT 55747
 159 159
 55748 ACTGGTTTGGGTAATTTGTTTGGGCAATGTTCTACATCTGATACA 55797
 159 159
 55798 GAGGTTTGAAGTACTTTACTAGTAGAGTATGATCTAGATGAAGAA 55847
 159 159
 55848 TGTTTGTTGTTTACTGTTTTCGATAGTTTGTAGTGTGTTGAATGTC 55897
 159 159
 55898 CAAAGCCGATGTTGTTATGATTTTGATATTGTTTATAGCAATG 55947
 160Leu.AspA 162
 55948 TTGCTACAGATGATTAACAGAGCTTCTTCTGTTTGAAGTATTTGATG 55997
 162 laProAlaLeuGlnAspAspPheTyLeuAsnLeuValAspTrpSerSer 178
 |||||.....
 55998 CACGGGCTTCAAGATGATTTTATCTGAATCTGTGATGATGCTGCA 56047

179 HisAsnValLeuAlaValGlyLeuGlyAsnCysValTyLeuTrpAsnAl 195
 :|||.....
 56048 CAAATGTTTCTAGCAGTGGAGTACAGGAACGTGTGTATTTATGGAATGC 56097
 195 acYserSerIys..... 199
 56098 TTTGATGAGCAAGAGTGAAGTATTTTGTGCTGCTGCAAAATTTCTAATAAT 56147
 199 199
 56148 CACAGTACTTCAAAATTTGTTTGGTGTGTTGATTTCTCTCTCAAAATC 56197
 200ValThrLysLeuCysAspLeuGlyValAlaAspAspCysValCysE 214
 |||||.....
 56198 CAACAGGTTACTAAGTATGATGATCTCGAGCTGAGATAGTGTGCTC 56247
 214 rValGlyTrpAlaGlnArgGlyThrHisLeuAlaValGlyThrAsnArg 231
 |||||.....
 56248 ACTGGTGGGGCTTACGTCGAACTCATCTGGCTGTGGAACCTAGTACCG 56297
 231 LysValGln.Ile..... 235
 56298 GGAAGTTGAGTTAGATCTCATTTACTCTTGTGTTTGTCTGTACTA 56347
 235 235
 56348 AGAATTTACGTTACTTTCGCACTTCTGATTTAGAAAAACATATTTTAA 56397
 235 235
 56398 GTTAACTATTTTACTTTACACATGATTTAGATTGACCTTATCTGTGNC 56447
 235 235
 56448 TGTAGGATTTGCCATATCTTTACTGCTATGATCTGATCATTTAGTGG 56497
 235 235
 56498 TATGATTTACAAAGATTTTCTTTATCGAAGAAATGCTCATTTA 56547
 235 235
 56548 AGCATTTACAGGCTTGTAAATTTTGAAGTGCCTGATGAAGCTA 56597
 235 235
 56598 AATCCTTGTGTTAAGATGAAGATCTGCATATAAACTTGGCGGA 56647
 235 235
 56648 TAAGATTTCAATATATCTGAAGAAATTTATGATGTTCTTTATGCAATAT 56697
 236TrpAspAlaAlaArgCysLysLys 243
 |||||.....
 56698 TCACCTTACTTGTGTTTGTGATATGAGATGCGTCAGCGCAAGAGA 56747
 244 IleArgSerMetGluGlyHisArgLeuArgValGlyAlaLeuAlaTrpSe 260
 |||||.....
 56748 ACAAGAACATGGAAGTCACTGCTCAAGAGTTGGAGCCCTGCAATGGGG 56797
 260 rSerSerLeuLeuSerSerGlyValArgAspLysAsnIleTyArgIleArgA 277
 :|||.....
 56798 TTTATCGGTTCTGTCTGTCGTGAGCAGACAAAGATATTCTTCAGAGAG 56847
 277 spIleArgThrGlnGluAspPheValSerLysLeuSerGlyHisLysSer 293
 |||||.....
 56848 ACATTAAGTGTCAAGAGATCATGTGATTAATTTGGCAGCATTAATCT 56897
 294 GluValCysGlyLeuLysTrpSerTyTrpAsnArgGluLeuAlaSerG 310
 |||||.....
 56898 GAGGTATGCGGACTCAAGTGTCTTATGACAAACAGAGACTAGCATCTGG 56947
 310 yGlyAsnAspAsnLys..... 315

56948 TGGAACGACAATAGGGTATGTACGATCAATTGATTGGCGTTAATTAG 56997
|||||
315 315
56998 TCCTCTGTTAAACCCCAACATAGCTAAAGTATTAATATTTCGAG 57047
316 LeuPheValTrpAsnGlnHisSerThrGlnProValLeuLysTyrCysG1 332
|||||
57048 CTTTGTGTATGGACCAACATCTCAACACACACGGGTTTGAATATATAGCA 57097
332 uHisThrAlaAlaValLysAlaIleAlaTrpSerProHisLeuHisGlyL 349
|||||
57098 ACACATGAGAGCTGTAAAGCCATTGCTGGTCTCTCATGTTCATGGGC 57147
349 euleuAlaSerGlyGlyThrAlaAspArgCysIleArgPheTrpAsn 365
|||||
57148 TTTTGTCTTCTGTGGTGCTGACTGCTGATGATGATGATGATTTTGGAT 57197
366 ThrTrpThrAsnSerHisLeuSerCysMetAspThrGlySer..... 379
|||||
57198 ACAACACGAAATCTCATTTAAGTTCATAGTACTTGCAGACGAGTAA 57247
379 379
57248 TTTCTTAAGTTCAGATGATGATGATGATGATGATGATGATGATGATG 57297
380GlnValCysAsnLeuValTrp 386
|||||
57298 TTAATATTCATCTTTGTCTTATATATTACAGATATGCAATCTAGCTGG 57347
387 SerLysAsnValAsnGlnLeuValSerThrHisGlyTyrSerGlnAsnG1 403
|||||
57348 TCTAAACAGTAAACGAGCTTGTAGCACACACGATCTCTCAGAACCA 57397
403 nIleIleValTrpArgTyrProThrMetSerLys..... 414
|||||
57398 AATCATGTCTGTAATATACCAACATGTCCAAAGTATGAATCTCTCTGA 57447
414 414
57448 CCTATATTTACTTCCATCATTTGAGACATCATGATTTATTCATGCTTT 57497
415LeuAlaThrLeuThrGlnHisThrTyrArgValLeu 426
|||||
57498 TTTTGTCTTGTACAGATTCCTACTTAACCGGTCCACATATCCAGATCTT 57547
426 uTyrLeuAlaIleSerProAspGlyGlnThrIleValThrGlyAlaGly 443
|||||
57548 ATATCTTGGCGTTTCAACCCGATGAGACAGATTTGTACAGACAGACAGAG 57597
443 spGluThrLeuArgPheTrpAsnValIlePheProSerProLysSer..... 457
|||||
57598 ATGAACCTTAAGTTCTGGAATGTTTCCCTCCCAAAATTCACAGTA 57647
457 457
57648 AAACCAACTACTACTTGTGTAATTACTTAATGATCAAAAATATGCT 57697
458GlnAsnThrG 461
|||||
57698 ATTTCAGAAAGAGCTGATTCGATGTTTGGCTGTGTTTCAGAACACGG 57747
461 LuSerGlnIleGlyAlaLeuSerLeuGlyArgThrThrIleArg 475
|||||
57748 ATATGTAATTCGGGTCTTCTTGTGTAGAACACAAATTCGG 57791
seq_name: gb_pl:ATT26M18
seq_documentation_block:
LOCUS ATT26M18 90824 bp DNA linear PLN 15-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone T26M18 (ESSA
project).

ACCESSION	AL078606
VERSION	AL078606.1 GI:5002514
KEYWORDS	
SOURCE	
ORGANISM	Arabidopsis thaliana thale cress. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE	1 (bases 38390 to 39150); 59320 to 59877; 60154 to 60701) Bevan,M., Volckaert,G., Gympfpretz,B., Voet,M., Robben,J., Bancroft,I., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 90824) Bevan,M., Hilbert,H., Braun,M., Holzer,E., Brandt,A., Duesterhoeft,A., Bancroft,I., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
AUTHORS	Unpublished
JOURNAL	EU Arabidopsis sequencing/project.
TITLE	Direct Submission
COMMENT	Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de,meyer@mips.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambidge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bsrc.ac.uk Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/ .
FEATURES	
source	1..90824 /organism="Arabidopsis thaliana" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4"
misc_feature	1..6968 /note="overlap to BAC T5C23: please refer to this entry for analysis and annotation" complement(7302..7490) /gene="AT4g11800" /number=1
gene	7302..12110 /gene="AT4g11800" complement(join(7302..7490,7664..7948,8032..8183, 8370..8667,8866..8967,9050..9382,9653..9852,9915..10138, 10220..10449,10560..10647,10762..10860,10991..11106, 11193..11441,11552..11777,11863..12110)) /gene="AT4g11800"
CDS	complement(join(7302..7490,7664..7948,8032..8183, 8370..8667,8866..8967,9050..9382,9653..9852,9915..10138, 10220..10449,10560..10647,10762..10860,10991..11106, 11193..11441,11552..11777,11863..12110)) /gene="AT4g11800" /note="similarity to other hypothetical proteins -Arabidopsis thaliana Contains Crystallins beta and gamma 'Greek key' motif signature AA488-503;Prokaryotic membrane lipoprotein lipid attachment site AAI54-164 contains EST gb:A1995638.1" /codon_start=1 /product="putative protein" /protein_id="CAB44318.1" /db_xref="GI:5002515" /translation="MESFRTLLHTYTPYPHSHRAITAVLFGCLFFISSDNMTLIE KFYSKWMNSACLGEFPYFSSPFIQKITRPVSNFSRMYIAMILVALYLPLNPSOM GLDRNMNLSELTIVTISLEFLVFHFIIIFLGVLGYGRVAGRPETILLONCAVL SMACCITFSGNRNAVLRQKRLGROYTSWFSFKREHNMTMLAFIMRNLIKQVS SWRPVSAASDIYPLSKWFITYGETIACNGSCPDADEISPTISLATFTGLYTANTVE RSTWALTHLP.LSVDKYEKLNNOLKPDFLDMPVWSGSADLFKVPDLGLVASVFLG RFEDRMLOAAMPKSGDASGRKKELVDHLNEKDVPFDPMADGDGNSYSVALILO PSLVPPANNFTSLPRGNVLTLIGDLAYNPSSFTFYERKLDFCFPEYALOPRWYNKS IAYVKPLPGVSDIKSYEGPOECFLIDRGHGGEFOVSAAPIQINMFPSLINFMYTCH KSWLGGMLMGPKSSTRALQIPKGMWTFGLDALHDGDIVDGKPFPSSELVTKRKVESDA


```

47 ThrIleTyrSerAspArgPheIleProSerAlaSerIysPheAl 63
76749 CCAATTACTCAGATATGGTTATTTCCAAAGATGATCTGTTCCAAATTCGC 76700
63 AleuPheAspIle.....AsnThrProThr.....GluGlyArgAspA 76
76699 GCTTTTCATTTGGCATCTCTTCGCCGAATATGAAGAAAGATGGAAAGAG 76650
76 spSerSerAlaTyrThrThrIleuValgThrAlaLeuPheGlyPro 92
76649 ATGGGGCTGGTTCTTATCCAGCTCTTTGAAACGGCGCTTTT..... 76606
93 AspValAlaGlyProValThrProGluLysThrAspSerProSerMetH 109
76605 .....GGTCGGGTGACGCGCGAGAAAGATGATGTTGTAATGGCTT 76565
109 rIleuProAsnArgAsnIlePheArgTyrIysThrGluThrArgGlnSerM 126
76564 TTCCTCGCGGGGAATTTTGGTTTAAAGCGGAAACGACGAGAGCTCTT 76515
126 eThIserIeuSerProPheMetAspAspAspPheValProGlyValAsn 142
76514 TGAATTTGATCCGCTTTT...GATTCGATGCTGTAGTGGCTTACG 76468
143 HisSerProValIysAlaProArgLysValProArgSerProTyrLys 158
76467 CCTAGTCTGTTTAAAGTCGCCGAGAAAGATCTTGGTCTCTTAAAGGT 76418
158 ..... 158
76417 CAGTGTGTTTGTTCATCTCGGGGGTTCCTTGTGGAGTTGAAATTAG 76368
158 ..... 158
76367 GTTGTAGATGTTGTTCTTGATGATTTGCTATTTGTTGGTAGATT 76318
158 ..... 158
76317 TAGGTTTGGAGCTGTGATTCGTGATGTTGTTGTCAATTCATCTA 76268
158 ..... 158
76267 TTTGTCATTTTANGATGGAACCTCTTGTAGTGTGGCTGATGTTCT 76218
158 ..... 158
76217 TTGAAACCTGTAGTGTGTCATTTTCTACTCTTATTCACGTT 76168
158 ..... 158
76167 TGAATTGAAATGTTTGTCTTTTGTGCATTTTGTCCGTGGATGAGCTGC 76118
159 ..... 159
76117 AAAAGCTCTTTTGTGTTTGTAGATTTGATGCGCGCTCTGCAAGATG 76068
168 spPheIleuAsnIleuValAspTyrSerSerHisAsnValIleuAlaVal 184
76067 ATTTTACTTGAAATCTCGTGAATGTCGGCAGCAAAATGTTCTTCGACG 76018
185 GlyLeuGlyAsnCysValTyrLeuTyrAsnAlaCysSerSerLys..Val 200
76017 GGATTAGGAACTGCTGTATTTTGTGGAAATGTTGATAGCAAGGTGAG 75968
200 ..... 200
75967 CTCTTTTCTTCTGAAATGTAATTTCTGTATCTCATTTCTTAGAT 75918
201 ..... 75918
75917 AGTGTCTTTTGATGAAATCTCTTCTGCTATTTGAGTAAGTAAGT 75868

```

```

203 euCysAspLeuGlyValAspAspCysValCysSerValGlyTyrAlaGln 219
75867 TATGCAATCTGGGGGTGATGAAAGCTGTTGCTCAGTGTGGTGGCACTA 75818
220 ArgGlyThrHisIleuAlaValGlyThrAsnAsnGlyLysVal..... 233
75817 CGTGGAAACAAATTTGGCTATTGGAAGTATGCGGAAACGATACAGCTTAG 75768
233 ..... 233
75767 AATATCAAGATTTGTTTGGCTATCTGAGTTATATCATGCTTATATATA 75718
233 ..... 233
75717 GTACTCTTGAATCTTGAATCGGAAGAAATCTTTTGGAAACTAGTT 75668
233 ..... 233
75667 GAATTAGACTTTGAGATGTTGCTTTTGCAAAATGTTATTCATCGAAT 75618
233 ..... 233
75617 GCATTGCTAGCTTACTTTCACCAATCATATGACTCATTAATGATC 75568
233 ..... 233
75567 GTTGGTTTGAAGATGATATTTGCTTAAACACCAACAGATCAAGTTAAT 75518
233 ..... 233
75517 GCAGCAATCTGATCTTATTAAGGTGATGATCTTATAGAGAA 75468
234 ..... 234
75467 CTTACGTTACTGATTTCTTAAATAGATATGAGATGCTTACAGTCA 75418
242 sLysIleArgSerMetGluGlyHisArgLeuAlaValGlyAlaLeuAla 259
75417 GAATATTGAACAATGAGAAAGTCAATCGCTAAAGTTGGAGCCCTGGCAT 75368
259 rPserSerIeuLeuSerSerGlyLysArgAspLysAsnIleTyrGln 275
75367 GGAGCTCAATGTTCTGCTTCAAGTAGTAGACAAAGACATCACTTACG 75318
276 ArgAspIleArgThrGlnGluAspPheValSerLysLeuSerGlyHisLys 292
75317 AGAGCATTTGCTACTCAAGAAGATCATGTCATTAACCTTAAAGTCAACA 75268
292 sSerGluValCysGlyLeuLysTyrPserTyrAspAsnArgGluLeuAla 309
75267 ATCCGAATCTGTGACTCAAAATGCTTCGCAATGCGAGCTAGCAT 75218
309 eArgLysIleAsnAspAsnLys..... 315
75217 CAGGTGAACGACAAATAGTAATATATATATCATCTTGTGTCTGA 75168
316 ..... 316
75167 AGTTAACATGAATTTTCAACATATGATGATTTCTAGATCTTGACGT 75118
316 upheValTyrPasnGlnHisSerThrGlnProValLeuLysTyrCysGlu 333
75117 GTTGTGATGAAACCAATTTCAACACAAACGCTTTTGAATCTGTGAAC 75068
333 sThrAlaAlaValAlaIleAlaTyrPserProHisIleuHisGlyLeu 349
75067 ACGCCGACAGAGTGAAACCATTCGTGCTCACATCTTTTGGACTT 75018
350 IeuAlaSerGlyGlyLysThrAlaAspArgCysIleArgPheTyrAsnThr 366
75017 CTGCTCTGTGTGCTGCTGCTGATAGATGATATCCGTTTCTGGAACAC 74968
366 rThrThrAsnSerHisLeuSerCysMetAspThrGlySerGlnVal..... 381

```



```

|||||.....|||||.....|||||
74967 AACGACAACTCATTTAAATGCGTAGATCCAAACAGCAAGTAATC 74918
381 ..... 381
74917 CGGTATGCTACTATTAACATCTGAGATTTCACATGTTGTTTATAG 74868
382 .....Cys..AsnLeuValTrpSerLys 388
74867 CAGCAATATCTTCTTCCACAGCAGAGTGATGTAATTTGGTATGCTAAG 74818
389 AsnValAsnGluLeuValSerThrHisGlyTrpSerGlnAsnGlnIleI 405
|||||.....|||||.....|||||
74817 AACGTGAATGACCTTGTGACACACAGGATTTCCAGAACCAATCAT 74768
405 eValTrpArgTrpProThrMetSerLys..... 414
|||||.....|||||.....|||||
74767 CCTCTGGAATATCCACACCATGCTTAAGTAAGATTTCTTCTCTACT 74718
414 ..... 414
74717 TCGTTTTTTTATGTGTCGATTTTGGATTTTGATGAGAACCTGTT 74668
415 .....LeuAlaThrLeuThrGlyHisThrTyArgValLeuTyrIle 428
74667 CAGTATATGTTGGCACTCTGCTGCTACCTGACCGCGTTGTTACT 74618
428 uAlaIleSerProAspGly.Gln..... 435
|||||.....|||||.....
74617 TCGGCGTCCAGCGAGATGACAGAGTACTTTTCTTCTCTCTGCTC 74568
435 ..... 435
74567 ATCTCTTAATCATCATCTTAATTTTCCCTACTTCTTAACACTTG 74518
436 .....ThrIleValThrGlyAlaG 442
74517 CATCAACACACTTCTTCTGATTTGTTGCAGCAATGTGACAGAGAGAG 74468
442 lYAspGluThrLeuArgPheTrpAsnValPheProSerProLys..Ser 457
74467 GAGATGAACCTGTGAGATTCTGAGATGCTTCCCTTCCGAAATCTGAG 74418
457 ..... 457
74417 GTAAAGAACCCGAGCATATTCGCCATTTCTAATCCCAATTTCTTGA 74368
457 ..... 457
74367 AATGACAACTGATTAACACACACTTCTCTTTTCTTTTATTTT 74318
458 .....GlnAsnThrGluSerGluIleGlyAlaLeuSerLeuGlyArgT 472
74317 CATCTCTGACAGCAGAGAGGAAATCGGGCATTAATCTTTGTAGAA 74268
472 hTThrIleArg 475
|||||.....
74267 CAACATCCGG 74257
seq_name: gb_pl:ATCHRIV33
seq_documentation_block: 190026 bp DNA linear, PLN 16-MAR-2000
LOCUS ATCHRIV33
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 33.
ACCESSION AL161533
VERSION AL161533.2 GI:7267889
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 2832 to 3389; 3666 to 4213)

```

```

AUTHORS Volckaert, G., Gijmenez, B., Voet, M., Robben, J., Mewes, H.W.,
JOURNAL Lemcke, K. and Mayer, K.F.X.
REFERENCE Unpublished
AUTHORS 2 (bases 123056 to 190026)
JOURNAL Peters, S.A., van Staveren, M., Dirkse, W., Stiekema, W., Mewes, H.W.,
AUTHORS Lemcke, K. and Mayer, K.F.X.
REFERENCE Unpublished
AUTHORS 3 (bases 1 to 34336; 23831 to 131430)
JOURNAL Hilbert, H., Braun, M., Holzer, E., Brandt, A., Duesterhoeft, A.,
AUTHORS Mewes, H.W., Lemcke, K. and Mayer, K.F.X.
REFERENCE Unpublished
AUTHORS 4 (bases 1 to 190026)
JOURNAL EU Arabidopsis sequencing project.
AUTHORS Direct Submission
TITLE Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
JOURNAL Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
biochemie@biochem.mpg.de, mayer@mps.biochem.mpg.de, proj/thal/
Coordinator: Mike Berran, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.berran@jic.ac.uk
COMMENT Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
this fragment has an overlap with ATCHRIV32 at the 5' end and an
overlap with ATCHRIV34 at the 3' end.
FEATURES
source Location/Qualifiers
1..190026
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
5417..6920
/gene="AT4g11890"
5417..5486
/gene="AT4g11890"
/number=1
join(5417..5486,5584..5675,5763..5973,6065..6302,
6396..6534,6612..6920)
/gene="AT4g11890"
/note="strong similarity to KI domain interacting kinase 1
-zee mays, PIR2:T02053
contains protein kinases signatures and profile AA159-171"
AA44-66:Protein kinases signatures and profile AA159-171"
/codon_start=1
/product="protein kinase-like protein"
/protein_id="CAB76232.1"
/db_xref="GI:7267890"
/translation="MAVTSLDIVFRRRRKKSTEFISVFEEDIDTIKAATNDSELY
GRGGRFYKRGKLNQGEIAYKILSTSTREPROPHNLLITSKKHNLNLGRCF
KRDQGLYERFNSLDFIUDPRAQOLWEMKRNITDGIARLRLHESGLWY
KRDIPGNILDSLDPKIVGFELARWQGEENAEETIEIVGTGYLPEIRSGRVS
KSDVYAGVTLITLISRRKASVGDSDLIYVRCWNRGAIVIHVMEERERS
ISEILRYHIALCVDENARRRPNIDKVIHMFSEFSPDPFTGNRLVDEETNMPW
SPSLRPHSSVSPISSR"
5487..5583
/gene="AT4g11890"
/number=1
5584..5675
/gene="AT4g11890"
/number=2
5676..5762
/gene="AT4g11890"
/number=2
5763..5973
/gene="AT4g11890"
/number=3
5974..6064
/gene="AT4g11890"
/number=3
6065..6302
/gene="AT4g11890"
/number=4
6303..6395

```



```

76  spSerSerAlaThrThrThrLeuLeuArgThrAlaLeuPheGlyPro  92
   |||.....|||.....|||.....|||.....|||.....
20161 ATGGGGCTGGTCTTATCGAGCTCTTTGAAAGCGCCCTTTT..... 20118
93  AspValAlaGlyProValThrProGluLysThrAspSerProSerMetIth  109
   |||.....|||.....|||.....|||.....|||.....
20117 .....GGTCCGGTGACCGCGGAGAAAGTGATGTTGTAATGGGCT 20077
109 rIeuProAsnArgAsnIlePheArgTyrLysThrGluThrArgGlnSerI  126
   |||.....|||.....|||.....|||.....|||.....
20076 TTCTCCGCGGGGAATATTTAGGTTTAAAGCGAAACGACGAGGCTCT 20027
126 eHisSerLeuSerProPheMetAspAspAspPheValProGluValAsn  142
   |||.....|||.....|||.....|||.....|||.....
20026 TGAATTTGATCCGCCCTTT...GATCTGATGCTAGTACGCTTAC  19980
143 HisSerProValLysAlaProArgLysValProArgSerProTyrLys..  158
   |||.....|||.....|||.....|||.....|||.....
19979 CCTAGCTCTGTAAGTCCGCGAGAAAGATCTTAGGCTCCTTAAAGCT  19930
158 .....  158
19929 CAGTGTGTTTGTGATCTTCGGGGGTTCTTGTGAGTTGAAATTAGG  19880
158 .....  158
19879 GTTGTAGATGTGTCTTCTTGATGATTTGTGATGTTGGAGAT  19830
158 .....  158
19829 TAGGTTTTGAGAGTTGATCTTGATGTTGTTGCATTCATTCANCTA  19780
158 .....  158
19779 TTTGGTCATTATGATGAACTCTTCTGCTAGTGGCGCTGATTTGTT  19730
158 .....  158
19729 TTGGAACCTGCTAGTGTGTTGTTGATTTTCTACTCTTATTCAGATT  19680
158 .....  158
19679 TGATTAGAAATGTTTGTCTTTGTCGCAATTTGCTGTGGAGAGTGG  19630
159 .....  159
19629 AAAAGCTTCTTTGTTTGTAGTATTTGATGCGCGGCTCTGCAAGATG  19580
168  spPheTyrLeuAsnLeuValAspTyrSerSerHisAsnValLeuAlaVal  184
   |||.....|||.....|||.....|||.....|||.....
19579 ATTTTACTTGAATCTCGATGCTGCGCACAAATGTTCTTGCAGTGG  19530
185 GlyLeuGlyAsnGlyValTyrLeuTyrPasnAlaCysSerSerLys..Val  200
   |||.....|||.....|||.....|||.....|||.....
19529 GGATTAGGAACTGCGTGTATTTGTGAATGCTGTAGTACCAAGGTAG  19480
200 .....  200
19479 CTCTTTTCTTCTGAATTTGTAATTTCTGATCTCATCTTCTTAGAT  19430
201 .....  201
19429 AGTGTCTTTTGTGATGGAATCTTTCTGCTATTTGCAAGTAAGTACTAGT  19380
203  eucCysAspLeuGlyValAspAspCysValCysSerValGlyTyrAlaGln  219
   |||.....|||.....|||.....|||.....|||.....
19379 TATGGCATTTGGGGGTGATGAACCTGTTTGTCTCAGTTGGTGGCACTA  19330
220  ArgGlyThrHisLeuAlaValGlyThrAsnAsnGlyLysVal.....  233
   |||.....|||.....|||.....|||.....|||.....
19329 CGTGAACACATTGCTATTTGAACTAGTAGCGGACACAGTACAGGTTAG  19280

```

```

233 .....  233
19279 ATATCAAGATTTGTTTGGCTATCTAGTATTATACATGCTTATATATA  19230
233 .....  233
19229 GTACCTCTTGACTCTTGAATTCGGAGAGAAATCTTTTCGAAACTAGGTT  19180
233 .....  233
19179 GAATTAGACTTTGAGAGATGCTCTTTTGCAAATGTTATTTCTAGCTGAA  19130
233 .....  233
19129 GCACTTCTAGCTTACTTACCGAATCATATGATGATTAAGTACTTATGAC  19080
233 .....  233
19079 GTTGGTTGTTAAGATGGAATTTGCTTAAACAACAAGTCAAGTTAAT  19030
233 .....  233
19029 GCAGCCATCTGATCTTATTAAGGTGTCATGATCTTCTTATGAGAA  18980
234 .....  234
18979 CTTACGTTTACTGATCTTCTTAAATAGATAGGAGATGTTACGATGCAA  18930
242  sLysIleArgSerMetGlnGlyHisArgLeuArgValGlyAlaLeuAla  259
   |||.....|||.....|||.....|||.....|||.....
18929 GAATATTAGAACAAATGGAAGTCAATCGCTAAGAGTTGGAGCCTTGCAAT  18880
259  rpSerSerLeuLeuSerSerGlyLysArgAspLysAsnIleTyrGln  275
   |||.....|||.....|||.....|||.....|||.....
18879 GGAGCTATCTGTTCTGCTCTTCAAGTAGTAGACAAAGACATACTTACG  18830
276  ArgAspIleArgThrGlnGluAspPheValSerLysLeuSerGlyHisLys  292
   |||.....|||.....|||.....|||.....|||.....
18829 AGAGACATTCGTACTCAAGAAATCATGTCATTAACCTTAAGCTGACAA  18780
292  sSerGluValCysGlyLeuLysTyrSerTyrAspAsnArgGlyIleuAla  309
   |||.....|||.....|||.....|||.....|||.....
18779 ATCCGAATCTGAGACTCAAAATGCTCTTGGACAAATCGGAGCTAGCAT  18730
309  eArgLysAsnAspAsnLys.....  315
18729 CAGGTGAAACGACAAATAGATAAATATATATATCATCATCTGTGCTGA  18680
316 .....  316
18679 AGTTAACATGAATTTTCAGCAATATGTAATGATTAGATCTTGCAGCT  18630
316  upPheValTyrAsnGlnHisSerThrGlnProValLeuLysTyrCysGln  333
   |||.....|||.....|||.....|||.....|||.....
18629 GTTTGTATGACACACATTTCAACACACACCGGTTTTCAGATCTGTGAAC  18580
333  IsThrAlaAlaValLysAlaIleAlaIleTyrSerProHisLeuHisGlyLeu  349
   |||.....|||.....|||.....|||.....|||.....
18579 ACGCCGACGACGTAAGCAATGCGGTGCTCCACATCATTTTGGACTT  18530
350  LeuAlaSerGlyGlyGlyThrAlaAspArgCysIleArgThrPasnThr  366
   |||.....|||.....|||.....|||.....|||.....
18529 CTTCCTCTCTGTGTGGCTGCTGCTGATGATGATGATGCTTTCGGAACAC  18480
366  rThrThrAsnSerHisLeuSerCysMetAspThrGlnGlySerGlnVal....  381
   |||.....|||.....|||.....|||.....|||.....
18479 AACGACAAACATCTATTTAAATTCGATGATACCAACAGCCAGGTAAATC  18430
381 .....  381
18429 CGGTATGCTAATTATTAACATCTGAGATTTCACATTTGGTTTATAG  18380
382 .....  382
Cys..AsnLeuValTyrSerLys  388

```

```

18379 CTGCATATCTTCTTCCACAGCAGTGTGAATTTGGTATGCTAAG 18330
389 AasnValAsnGluLeuValSerThrHisGlyTyrSerGlnAsnGlnIleI 405
18329 AACGTAAAGACGTGTGTAGACACACGGGTATTCACCAACCAATCAT 18280
405 eValTPATgTyrProThrMetSerLys..... 414
18279 CGCTGGAATATCCACCATGTCTAAGTAATCTTCTCTTACT 18230
414 ..... 414
18229 TCGTTTTTTTATGTGCTGTAATTTGGATTTTGATGAGACCTGTT 18180
415 ..... LeuAlaThrLeuThrGlyHisThrTyrArgValLeuTyrIle 428
18179 CTGTATTAAGTTGGCACTCTGCTACTGCTACCTGACCGGCTGTACT 18130
428 uAlaIleSerProAspGly.Gln..... 435
18129 TCGGGGTACCGGAGTGGAGAGAGGTACTTTTCTTCTCTGCTGCTC 18080
435 ..... 435
18079 ATTCTTCAATCATCATCTTATATTTTGGCTACTTCTTATACACTTG 18030
436 ..... ThrIleValThrGlyAlaG 442
18029 CATCAACACACTTCTTCTTATTTGTCGACAAATTTGGACAGAGACAG 17980
442 lYAspGluThrLeuArgPheThrPasnValPheProSerProLys..Ser. 457
17979 GAGATGAACCTTGAAGATTCTGAAATGCTTCCCTTCCGAAATCTCAG 17930
457 ..... 457
17929 GTAAAGAACCCGAGACATATTCGCCATTTCTAATCTCAATTTCTTGA 17880
457 ..... 457
17879 AATGAAACATGTATTAACACACAACTTCTTTCTTTATTTT 17830
458 ..... GlnAsnThrGluSerGluIleGlyAlaLeuSerLeuGlyArgT 472
17829 CATCTCTGAGACGAGGAGAGGAATCGGGCATTTCTTTGGTACAA 17780
472 hTThrIleArg 475
17779 CAACAATCCGG 17769
seq_name: gb_BC013413
seq_documentation_block:
LOCUS BC013413 1866 bp mRNA linear PRI 04-SEP-2001
DEFINITION Homo sapiens, clone MGC:4589 IMAGE:3160334, mRNA, complete cds.
ACCESSION BC013413
VERSION BC013413.1 GI:15426580
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
REFERENCE
AUTHORS Strausberg, R.
JOURNAL Direct Submission
Submitted (31-AUG-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
NTI-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgaps-femail.nih.gov

```

```

FEATURES
source
1..1866
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:4589 IMAGE:3160334"
/tissue_type="Brain, neuroblastoma"
/clone_lib="NIH_MGC_19"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
199..1680
/codon_start=1
/product="Unknown (protein for MGC:4589)"
/protein_id="AAH13413.1"
/db_xref="gi:15426581"
/translation="MDQYERLRLOIYQIENMPRTYERRLTPASSPVSPSKH
GREFIPSRAGAWSVNFRINENKSPQNRKADATSDNKKDLAVALKNELDA
GLEKQVDPQTEDERLRQSTPEPKGLFTYLSLTKRSPPDGDVSPVSNKSLQ
LSPRRPKRISKIPKVLDAPELDQDLYLNVMSLNVSLVGLCYLYMSACTSQ
VRLCDLVEGDSVTSVGSERGNVAVGTGKGVQIWDAAAGKLSLSEHTARVGA
LAWNAEOLSSGSRDMLQDRITRPDQSEERLGGHGVNCGGLASGGGTRDRIRNTITGP
NKLIVMNSSLSPVQYTEHLAAKAKIMSRHOGGLASGGGTRDRIRNTITGP
LQCIDTGSQVNCNLAWSKHANLVSHTGISQNIIVMYRPSLTQYAKLGHSTRLYLA
MSPDEALVTGAGDETLRFNWFVSKTRTSKESVSLNLTFRIR"
BASE COUNT 379 a 601 c 595 g 291 t
ORIGIN
alignment_scores:
Quality: 1294.00 Length: 495
Ratio: 3.469 Gaps: 14
Percent Similarity: 75.354 Percent Identity: 54.343
alignment_block:
US-09-701-572-2 x BC013413 ..
Align seg 1/1 to: BC013413 from: 1 to: 1866
6 AasnArgAsnProProThrSerThrValArgAspAsnSerProProPr 22
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
247 AATGAAACACGATGCCACCGCTCACAGAGATCGGGGACCTGACGCC 286
22 GlnProSerProGluSerLeuArgHisValSerArgMetIleAsnSerA 39
297 TGCACGCTCCCAAGTTC..... 315
39 snHisTyrThrSerProSerArgThrIleTyrSerAspArgPheIlePro 55
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
316 .....TCGCCCGACAG.....CACGAAACCGCTTCATCCCC 348
56 SerArgSerAlaSerLysPheAlaLeu.....PheAspIleAsn..... 68
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
349 TCCAGAGCCGAGAGCCACACTGAGCTGACACTTCACAGATTTACGAGA 398
69 .....ThrProGluGlyArg.....AspAspSerSerSer..... 79
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
399 TGAGAACTCTCCAGTACAGACGGAAGCCAGGACGCGCACCCTAGACA 448
80 .....AlaTyrThrThrLeuLeuArgThrAlaLeuPhe 90
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

449 ACGGCAAGAGAGCGCTGACCTACTGTGCTCCGTCAAGAAATGAGCTGCTG 498

91 G1Y.....ProAspValAlaIglProValThrProGluIysThrAs 104

499 GGTGCCCGGCATCAGAGAGCTCAGAGACCCCGGAGCTAGGACCCCGAGCT 548

104 pserProsermetThrLeuProAsnArgAsnIlePheArgIyrIysThrG 121

549 GCAAGCCCTCC..ACGCTGAGAAAGAAAGGGTCTGTCACTACGATTCCTTA 595

121 IuThrArgGlnSerMet.....HisSer 128

596 GCACCAAGGGCTCCAGCCCGCATGACGCGACAGATGATGTCTCCTCACTCC 645

129 LeuSerProPheMetAspAspAspPheValProGluValAsnHisSerP 145

646 CTGTCTCCCGTCAGCAGCAAGAGAGC....CAGAAGCTGCTCCGGTCCCC 689

145 oValIysAlaProAlaGlyValProArgSerProTyrIysValIleuAspA 162

690 CCGGAAGACCCACCGCAAGATCTCCAAATATCCCTTCAAGGTGCTGACG 739

162 IAProAlaIleuGlnAspAspPheTyrIleuAsnIleuValAspTrpSer 178

740 GCGCCGAGCTGAGGAGACACTTATCACTCAATCTGATGAGTGTGCTGCTCC 789

179 HisAsnValIleuAlaValIglIleuGlyAsnGlyValIyrIleuTrpAsn 195

790 CTCAATGTGCTCAGCGGTGGGCTAGGACACCTGCTGCTGAGAGTGC 839

195 acySerSerIysValIThrIysIleuCyAspIleuGlyAlaasp..Asp 211

840 CTGTACACAGCCAGGAGTACGCGGCTGTGACCTCTCAATGGAAAGGGGCT 889

211 ysaValCySerValIglYrTrpAlaGlnArgIglYrHisIleuAlaValIgl 227

890 CAGTGACCTCCGTGGGCTGTGCTGTAGCGGGGGAACCTGTGGCGGTGGCG 939

228 ThrAsnAsnGlyLysValAlaGlnIleTrpAspAlaIleArgCysIysI 244

940 ACACACAAAGGGCTTGCTGCTCAGATCTGGGAGCAGCCGAGGGAAGAACT 989

244 eArgSerMetGluGlnIHisArgIleuArgValIglValIleuAlaTrpSer 261

990 GTTCATGTTGGAGGGCCACACGCGCAGCGGTGGGGGCTGGCTGGAAATG 1039

261 erSerIleuSerSerArgIglYrArgAspIysAsnIleTyrGlnIArgAsp 277

1040 CTGACAGAGTGTCTGCTCGGGAGCCCGACCCGATGATCTTCGACAGAGGAC 1089

278 IleArgThr.....GlnGluAspPheValSerIysIleuSerGlyHis 291

1090 ATCCGACACCCCGCCACTGCAGCTCGAG..CGCGCTGTGACAGGGGCCA 1133

291 slySerArgIuValIcysGlyIleuIysTrpSerTyrAspGlnArgIuLeuA 308

1134 CCGGAGAGGGGTGCTCGGGCTCAATGTGTCCACACACACAGCACTCTCG 1183

308 IAserGlyValAsnAspAsnIysIleuPheValTrpAsnGlnHisSerThr 324

1184 CTTGGGGGGGCAACACACAAAGCTGCTGTGTGGAATCATCTCGAGCTG 1233

325 GlnProValIleuIysTyrCysGlnHisThrAlaIleValIysAlaIleAl 341

1234 AGCCCGCTGCAGACAGTACAGAGACACCTGGGGGCGTGAAGGCATATGCG 1283

341 atTrpSerProHisIleuHisGlyIleuIleuAlaSerGlyIglYrThrAla 358

1284 CTGTGTCCCAATCAGACAGGGCTGCTGTGCTCGGGGGGCGGACAGCTG 1333

358 spArgIysIleArgPheTrpAsnThrThrHisAsnHisIleuSerIys 374

1334 ACCGCTGATCCGCTTCTGGAACAGCTGACAGGAAACCACTATGAGATG 1383

```

375 MetAspThrGlySerGlnValCysAsnLeuValTrpSerLysAsnValAs 391
      ::::::::::::::::::::|
1384 ATCGACACGGGGCTCCCAAGTGTGCATTCGGCTGTCCAAAGCACCCAA 1433
      |:::|||||||
391 nGluLeuValSerThrHisGlyTyTrSerGlnAsnGlnIleIleValAlrPA 408
      |||||
1434 CGAGCGGTGTGAGACCACCGAGGGCTACTACAGAACAACGATCCTGTGTGA 1483
      |||||
408 rgtTrpProThrmMetSerLysLeuAlaThrLeuThrGlyHisThrTYRaaG 424
      :|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1484 AGTACCCTTCCTCGAACCCAGTGCGCCAACTGACCCGGGACATCCTACCC 1533
      |
425 ValLeuTyrrLeuAlaIleSerProAspGlyInThrIleValThrGlyAl 441
      |||||
1534 GTGCTGTACTCTGGCAATGTCCCTGATGGGAGGCCATGTGTACTGTGTGC 1583
      |
441 acgAAspGlyThrLeuArgPheTrypsinaValPheProSerProLysSerg 458
      |||||
1584 TGGAGACGAGACCCCTGAGGTCTGGAACGCTTT.....AGCAAAAACC 1627
      ||:::|
458 InasnThrGluSerGluIleGlyAlaLeuSerLeu 469
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1628 GTTCGACAAAGAGAGTGTGTGTCTGTCTCAACCTC 1662

seq_name: gb_pr:AF083810
```

LOCUS	AF083810	2991 bp	mRNA	linear	PRI 01-SEP-1999
DEFINITION	Human sapiens fizzy-related protein (FZR) mRNA, complete cds.				
ACCESSION	AF083810				
VERSION	AF083810.1	GI:5813826			
KEYWORDS	human.				
SOURCE	human.				
ORGANISM	Human sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
JOURNAL	1 (bases 1 to 2991)				
REFERENCE	Jin,D.-Y. and Jeang,K.-T.				
TITLE	Characterization of human fizzy-related protein				
REFERENCE	Unpublished				
TITLE	2 (bases 1 to 2991)				
REFERENCE	Jin,D.-Y.				
TITLE	Direct Submission				
REFERENCE	Submitted (12-AUG-1998) LHM, NIAID/NIH, 9000 Rockville Pike, Bethesda, MD 20892-0460, USA				
JOURNAL	Location/Qualifiers				
FEATURES	1..2991				
SOURCE	/organism="Homo sapiens"				
gene	/db_xref="taxon:9606"				
CDS	1..2991				
	/gene="FZR"				
	140..1621				
	/gene="FZR"				
	//function="cell cycle regulator"				
	//codon_start=1				
	//product="fizzy-related protein"				
	/protein_id="AAD52030.1"				
	/db_xref="GI:5813827"				
	/translation="WDDEYERRLRLQIYIÖNENMPRTYEMRRLTPASPVSYSKSKK				
	GDRIPSPRAGANWSVNFHRIINENKSPÖNRKADATSDNGKDLAYSALIKNELGLG				
	GIERVODPÖTDRRLQPESTPEKGLFTYSLSTRKSPDGDVDSYSLSPVSNKSÖK				
	LRSRPKIPRKISKIPEFKVLAPELÖDDYLLNLVWSSLNANVSGIGTCVYLMSACTS				
	VTYRLCDSVEGDSVTSVGMSEKGNLVAVGTGKGFQIWDAAAGKLSMLEGHTFARVAG				
	LAMNAEQLSSGSDRMILQDRITRPPLDSERLQGHQREVCGLKSTSHOLLASGNN				
	NKLIVNRRSSLSIPVOQYTEHLAAKATMSPHQHGILLASGGGTADRCIRIPNTLNGD				
	LÖCIDTGSÖGNTLMAKSHANVELSTHGISÖQILYWKYPSLTQVAKRLGHYRIYLYL				
	MSPGEAIVTGAAGETLRFNWFSEKSTRKESVYLNLFTRIR"				
BASE COUNT	570 a 991 c 920 g 510 t				
ORIGIN					
alignment_scores:					

Quality: 1294.00 Length: 495
 Ratio: 3.469 Gaps: 14
 Percent Similarity: 75.354 Percent Identity: 54.343

alignment_block:
 US-09-701-572-2 x AF083810 ..

Align seg 1/1 to: AF083810 from: 1 to: 2991

```

6 AsnArgSerProProProThrSerThrValAlaGAspAsnSerProPro 22
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
188 AATGAGAAACAGATGCCACGCGTCACAGAGATCGCGGACCTGACGCC 237
22 ocgUpSerProgluSerleuArgHisIvalSerArgMetIleAsnSer 39
|  |||||  |||  |||  |||  |||  |||  |||  |||  |||
238 TGCACAGCTCCCAAGTGTCC..... 256
39 snHisIyThrSerProSerArgThrIleTySerAspArgPheIlePro 55
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
257 .....TGCACAGCAAG.....CAGGAGACCGCTTCATCCCC 289
56 SerArgSerAlaSerIysPheAlaLeu.....PheAspIleAsn..... 68
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
290 TCCAGAGCCGAGCCACTGAGCGTGAACCTCCACAGAGATTACAGAGA 339
69 .....ThrProThrIleGlyArg.....AspAspSerSer..... 79
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
340 TGAGAAGTCTCCACAGTACAGAACCGGAAAGCCAGACCGCACCCTAGA 389
80 .....AlaTyThrThrIleuLeuArgThrAlaLeuPhe 90
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
390 ACGGAAAGACGCGCTGCGCTACTCTGCTGCTCAAGAAATGAGCTGTG 439
91 Gly.....ProAspValAlaGlyProValThrProGluLysThrAs 104
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
440 GGTGGCGGATGAGAGAGTGCAGAGACCCGCACTGAGAGCCGAGCT 489
104 pSerProSerMetThrIleuProAsnArgAsnIlePheArgIyIleThrG 121
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
490 GCAGCCCTCC...ACGCTGAGAGAGAGAGGTGTGTTCAGATATCCCTTA 536
121 IuThrArgIleSerMet.....Hiser 128
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
537 GCACAGAGCGCTCAGCCCGATGAGCGCAAGATGTCTCCCTACCTCC 586
129 IeuSerProPheMetAspAspAspPheValProGlyValAsnHisSerPr 145
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
587 CTGTCTCCCGTACGACACAGAGC.....CAGAAAGCTGCTCCGCTCCC 630
145 oValIysAlaProArgLysValProArgSerProTyIlyValLeuAspA 162
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
631 CGGAAACCCACCCGCAAGATCTCAAGATCCCTTCAAGGTGCTGAGCG 680
162 IAProAlaLeuGlnAspAspPheTyIleuAsnLeuValAspTyIleSer 178
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
661 CGCCGAGAGTGCAGGAGCACTTCACTCAATCTGTGTGACTGTGCTGCC 730
179 HisAsnValLeuAlaValGlyLeuGlyAsnCyValTyIleuTPAsnAl 195
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
731 CTCATATGTCTCAGCGTGGGTAGGACACCTGCTGACCTGTGAGAGCC 780
195 acySerSerIyValIleuLysLeuCyAspLeuGlyValAsp...AspC 211
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
781 CTGTACCAAGCAGGTGACGCGCTGTGACCTCTCACTGAGAGGCGACT 830
211 yValCySerValGlyTrpAlaGlnArgGlyThrHisLeuAlaValGly 227
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
831 CAGTACCTCCGTGGGTGTGTGAGCGGGAACCTGTGTGCGGTGGGC 880
228 ThrAsnArgIyLysValGlnIleuTrpAspAlaIleArgCysLysIle 244
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
881 ACACACAAAGGCTTGTGCTGAGTCTGGAGCGACGCGCAGGAGAAAGCT 930

```

```

244 eArgSerMetGluGlyHisArgLeuArgValGlyAlaLeuAlaTrpSerS 261
:  |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
931 GTCCATGTGTGAGGGCCACAGCAGCCGCTCGGGCCCTGGCATG 980
261 erSerLeuSerSerGlyIyArgAspLysAsnIleTyIleGlnArgAsp 277
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
981 CTGACAGACTGTCTGCTCCGGAGCCGCGACCCGATGATCTGACAGAGGAC 1030
278 IleArgThr.....GlnIleAspPheValSerIySerGlyHis 291
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1031 ATCCGACCCCGCCACTGCTGAGAG.....CGGCGCTGACAGGCCA 1074
291 slySerGluValCySlyLeuLysTrpSerTyIyAspAsnArgGluLeuA 308
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1075 CCGGACAGAGGTGTGGGGCTCAAGTGTCTCACAGACACCACTCTCTCG 1124
308 IAserGlyGlyAsnAspAsnLysLeuPheValIyTrpAsnGlnHisSerThr 324
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1125 CCTCGGGGGGACAGACACACACTGCTGCTGATCACTGAGGCTG 1174
325 GlnProValLeuLysTyIyGlyHisThrAlaValAlaValIleAla 341
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1175 AGCCCGCTGACAGCACTACAGAGACCTGCGCGCTGAAAGCCATCGC 1224
341 atPTrSerProHisLeuHisGlyLeuLeuAlaSerGlyGlyIyThrAlaA 358
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1225 CTGTCTCCCATCATCAGACAGGGGTGCTGCGCTGCGGGGGGACACAGCTG 1274
358 spArgCysIleArgPheTrpAsnThrThrTrpAsnSerHisLeuSerCys 374
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1275 ACCGCTGTATCCGCTTGTGAGACAGCTGACAGAGACAACTGCACTG 1324
375 MetAspThrGlySerGlnValCySAsnLeuValIyTrpSerLysAsnAla 391
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1325 ATGCACACGGGCTCCCAAGTGTGCAATCTGCTGCTGCTGCAACAGCCCA 1374
391 nGluLeuValSerThrHisGlyTyIySerGlnAsnGlnIleIleValIyTPA 408
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1375 CGACCTGTGAGACAGCAGCGGTACTACAGAACACAGATCTGTCTGGA 1424
408 rGlyTrpProThrMetSerLysLeuAlaIleThrLeuThrGlyHisIleTyIyArg 424
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1425 AGTACCCCTCCCTGACCCAGAGTGGCCAAAGCTGACCGGGCCTCTACCCG 1474
425 ValIeuTyIleuAlaIleSerProAspGlyGlyThrIleValThrGlyAl 441
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1475 GTGCTGTACTGTGCAATGTCTCCCTGATGGGAGGCCATCTCTGTGTGC 1524
441 agIyAspGluThrIleuArgPheTrpAsnValPheProSerProLysSerG 458
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1525 TGGAGAGAGAACCTTGAGTGTGGAACGCTTT.....ACCAAAACCC 1568
458 IAsnThrGluSerGluIleGlyAlaLeuSerLeu 469
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1569 GTTCGACAAAGAGATGTGTGTGTGCTCAACCTC 1603
seq_name: gb_pr:AB033068
seq_documentation_block:
LOCUS AB033068 5008 bp mRNA linear PRI 11-NOV-1995
DEFINITION Homo sapiens mRNA for KIAA1242 protein, partial cds.
ACCESSION AB033068
VERSION AB033068.1 GI:6330804
KEYWORDS
SOURCE
  Homo sapiens brain cDNA to mRNA, clone_lib:pbiscriptII SK plus
  Clone:fh10406.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
  1 (sites) Nagase,T., Ishikawa,K., Kikuno,R., Hirosewa,M., Nomura,N. and
  Ohara,O.
  Prediction of the coding sequences of unidentified human genes. XV.
  TITLE

```

The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro
DNA Res. 6 (5), 337-345 (1999)

JOURNAL
MEDLINE
20039619
REFERENCE
2 (bases 1 to 5008)
AUTHORS
Ohara, O., Nagase, T. and Kikuno, R.
TITLE
Direct Submission
SUBMITTER
Laboratory of DNA Technology, 1532-3 Yama, Kisarazu, Chiba
292-0812, Japan (E-mail: cdnainfo@kazusa.or.jp,
URL: http://www.kazusa.or.jp/huge/, Tel: +81-438-52-3913,
Fax: +81-438-52-3914)

FEATURES

source
1. 5008
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Fh10406"
/issue_type="brain"
/clone_lib="pbluescriptII SK plus"
1. 1516
/gene="KIAA1242"
1. 1516
/gene="KIAA1242"
/note="Start codon is not identified."
/codon_start=2
/product="KIAA1242 protein"
/protein_id="BA86556.1"
/db_xref="GI:6330805"
/translation="LTLPRAPCLAMDQDERLLRQIVIONENTMPRTVETMRRLTP
ASSPVSPPKSGDRFIPSRAGNMSVNFHRIENENKSPQNRKADATDNGKDLAY
SALKRELKAGIEKVEDQDPTDRRLQSPKGLFTYSLTKSSPDGNDVSPYS
LSYNSKSKLRLSPKPRKTSKIDPEKLDAPLDDPDLNLVDMSSLVNLSVGLGT
CYLWSACTSVYTRLCDLSVEGDSVTSVSGSERGNLVANGTKGVQITDAAGKIS
MEGHARVAGALAMNAEQSLSSRDMLQRIPTPLQSERLQGHROEVGLMNST
DHQILASGNDKLLVNHSSLSPPVOYTEHLAAVLAVALMSRHQGLASGGGTADRC
IREMVLTCGLOPLOCIDTGSQVCMIAWSKHANETVSHGSONQIIVWKYPSLTQVAKL
TGHSYRVILVAMPDEALVTGAGDETLRFNMVFSKTRSTKESVSLNLFTRIR"

BASE COUNT 939 a 1492 c 1680 g 897 t
ORIGIN

alignment_scores:
Quality: 1294.00 Length: 495
Ratio: 3.469 Gaps: 14
Percent Similarity: 75.354 Percent Identity: 54.343

alignment_block:
US-09-701-572-2 x AB033068 ..

Align seg 1/1 to: AB033068 from: 1 to: 5008

6 AsnArgAsnProProTherTherValArgAspAsnSerProPro 22
|||||
83 AATGAGAACACGATGCGACGCGTACAGAGATGCGACCGCTGACGCC 132
22 ocluproserProgluSerleuArgHisValSerArgMetIleAsnSera 39
|
133 TCCCAAGCTCCCAAGTGTCC..... 151
39 snHisTyrTherSerProserArgThrIleTyrSerAspArgPheIlePro 55
|||||
152TCGCCACGACAG.....CAGGAGACCGCTTCAATCCCC 184
56 SerArgSerAlaSerIlePheAlaLeu.....PheAspIleAsn..... 68
|||||
185 TCCAGAGCCGAGCCAACTGAGCGGAACTTCCACAGGATTAAAGAGAA 234
69ThrProThrGluGlyArg.....AspAspSerSer..... 79
235 TGAAGAGTCTCCAGTCAAGACCGGAAAGCCAAAGGCGCCACTCAAGACA 284
80AlaTyrThrThrLeuLeuArgThrAlaLeuPhe 90
|||||

285 ACGGCAAAAGACGCGCTGACCTACTCTGCCCTGCTCAAGATGAGCTGCTG 334
91 Gly.....ProAspValAlaGlyProValThrProGluTyrThrAs 104
335 GGGGCCGCGATGAGAGAGTGCAGAGACCCGACACTGAGAGACCGAGGT 384
104 pSerProserMetThrLeuProAsnArgAsnIlePheArgTyrTyrThrG 121
385 GCAAGCCCTC...ACGCTGAGAGAAAGAGGTCTGTACAGTATCCCTTA 431
121 LuThrArgIleSerMet.....HisSer 128
432 GCACCAAGGCTCCAGCCCGATGACGACGACAGAGTGTCTCCCTACTTC 481
129 LeuSerProPheMetAspAspAspPheValProGlyValAsnHisSerPr 145
482 CTGTCTCCGCTGACCAAGAGAC.....CAGAACCTCTCCGCGTCCCC 525
145 oValAlaProArgLysValProArgSerProTyrLysValLeuAspA 162
526 CCGGAAGACCCACCCGCAAGATCTCCAAAGATCCCTTCAAGGTGCTGAGC 575
162 laProAlaLeuGlnAspAspPheTyrLeuAsnLeuValAspTrpSerSer 178
576 CGCCGAGCTGCGAGAGAGACTTCTACTCTCAATCTGTGAGACTGTGCTCC 625
179 HisAsnValLeuAlaValGlyLeuGlyAsnValTyrLeuTrpAsnAl 195
626 CTCAATGCTGCTAGCGTGGGCTGAGGACCTGCTGCTACTCTGAGAGTGC 675
195 acYSerSerLysValThrLysLeuCysAspLeuGlyValAsp...AspC 211
676 CTGACACAGCCAGGTGACGGGCTGTGTGACTGTGACTGTGAGGAGGAGCT 725
211 yValCysSerValGlyTyrAlaGlnArgGlyThrHisLeuAlaValGly 227
726 CAATGACCTCTGCTGGTGTGTGAGCGGGGGAACCTGTGGGGGTGGG 775
228 ThrAsnAsnGlyLysValGlnIleTrpAspAlaAlaArgCysLysLysI 244
776 ACACACAAAGGCTTCTGTCAGATCTGGAGACGACCCGCAAGGAAGCT 825
244 eArgSerMetGluGlyHisArgLeuArgValGlyAlaLeuAlaTrpSerS 261
826 GTCCATGTTGAGAGGGCCACAGCGACCGCTCGGGGCGCTGCTGGAATG 875
261 eSerLeuLeuSerSerGlyLysArgAspLysAsnIleTyrGlnArgAsp 277
876 CTGAGCAGCTGTGCTCGGGAGCGGACCGCATGATCTGCAAGAGGAG 925
278 IleArgThr.....GlnGluAspPheValSerLysLeuSerGlyH 291
926 ATCCGACACCCCGCCACTGCACTGCGAG.....CGGGGCTGCAAGGCCA 969
291 sLysSerGlnValCysGlyLeuLysTrpSerTyrAspAsnArgGlyLeuA 308
970 CCGGACAGAGGTGTGCGGCTCAAGTGTGTCACAGACACACACTCTCTCG 1019
308 laSerGlyGlyAsnAspAsnLysLeuPheValTrpAsnGlnHisSerThr 324
1020 CCTCGGGGGGACAGCAACAGCTGTGTGCTGGAATCACTCGAGGCTG 1069
325 GlnProValLeuLysTyrCysGlnHisThrAlaAlaValLysAlaIleAl 341
1070 AGCCCGCTGACAGAGTACAGGAGACCTGCGGCGCTGAAGGCCATCGC 1119
341 aTrpSerProHisLeuHisGlyLeuLeuAlaSerGlyGlyTyrAla 358
1120 CTGCTCCCAATCAGACAGCGGCTGTGCTCGCTCGGGGCGGACACAGCTG 1169
358 spaArgCysIleArgPheTrpAsnThrThrAsnSerHisLeuSerCys 374
1170 ACCGCTGTATCCGCTTGTGAAACAGCTGACAGGACAAACCACTGCACTGT 1219

```

375 MetAspThrGlySerGlnValCysAsnLeuValTrpSerLysAsnValas 391
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1220 ATGACACAGGGGCTCCCAAGGTGCTATGCGCTGCTCCAGACGCCCA 1269
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
391 ngLueValSerThrHisGlyTrpSerGlnAsnGlnIleValTrpA 408
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1270 CGAGCTGCTGAGCAGCAGCGCTACTACAGAACAGATCTTGCTGGA 1319
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
408 rGlyTrpProMetSerLysLeuValaThrLeuThrGlyHisThrArg 424
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1320 AGTACCCCTCCCTGACCCAGGTGACAGTACGCGGACGCTACCC 1369
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
425 ValLeuTrpLeuAlaIleSerProAspGlyGlnThrIleValThrGlyAl 441
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1370 GTGCTGCTACCTGACATGCTCCCTGATGGGAGGCCATGCTCAGCTGTC 1419
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
441 agLysAspGlyThrLeuArgPheTrpAsnValPheProSerProLysSerG 458
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1420 TGGAGACGAGACCTTGAGGTCTGGAACGCTTT.....AGCAAAACC 1463
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
458 InAsnThrGlySerGlnIleGlyAlaLeuSerLeu 469
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1464 GTTCCGACAAAGAGCTGTGTCTGTCTGCTACCTC 1498
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::

seq_name: gb_AB013462

locus documentation_block:
LOCUS AB013462 1491 bp mRNA linear PRI 23-NOV-1999
DEFINITION Homo sapiens mRNA for Fzrl, complete cds.
ACCESSION AB013462
VERSION AB013462.1 GI:6463678
KEYWORDS Fzrl; Fzrl.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1491)
AUTHORS Kotani,S., Oyama,T. and Todokoro,K.
TITLE Human homologue of Fizzy-related protein
JOURNAL Published Only in Database (1999) In press
REFERENCE
2 (bases 1 to 1491)
AUTHORS Kotani,S., Oyama,T. and Todokoro,K.
TITLE Direct Submission
JOURNAL Submitted (29-APR-1998) Kazuo Todokoro, The Institute of Physical
and Chemical Research (RIKEN), Molecular Cell Science Laboratory;
3-1-1,Koyada, Tsukuba, Ibaraki 305-0074, Japan
(E-mail:todokoro@rct.riken.go.jp, Tel:81-298-36-9075,
Fax:81-298-36-9090)

FEATURES
Source
1..1491
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1491
/feature="fzrl"
1..1491
/feature="fzrl"
/feature="Fizzy-related protein"
/feature="Fzrl"
/codon_start=1
/product="Fzrl"
/protein_id="BA863679"
/db_xref="GI:6463679"
/translation="MDQYERLLNQIVIONETMRYTEMRTLTTPASSPVSSPSKH
GRFIPSRAGAMSVNFRINENKSPSONRKAQKATSGNGDGLAYSLAKNELIGA
GIEKVDQPTEDRLQSPTEPKGLFTYSLSTRSSPDGNDGVSLSPPVNSGOKL
LSRPRKTRIKSIKIPKVLADPELODDFVNLVMSLSNLVSVIGCTCVYLSMSACTSO
VTRICDLSVSGSVSVSGSERGNLYAVGTHTGCPVQINDAAGKRLSMLEGHATGVA
LAMNAEQLSSGSRDRILTDRLTRTPLOSERLQGHROEVGLKSTHHLIASGND
NKLVWNHSSLSPVQOYTEHLAVALAIAVSPHOGILASGGGTADRCIRFWMTLGGP
LOCIDTGSQVCMIAWSKANLELVSTHGYSQNDILVMKYPSILOVAKILGHSHYVLYLA
MSPDEAIIYTGAGDETLRFWNVFSKTRKTVKMSVSVNLFTLRIR"

BASE COUNT 312 a 479 c 452 g 248 t
ORIGIN

```

```

alignment_scores:
Quality: 1286.50 Length: 496
Ratio: 3.458 Gaps: 14
Percent Similarity: 75.000 Percent Identity: 54.234

alignment_block:
US-09-701-572-2 x AB013462 ..

Align seg 1/1 to: AB013462 from: 1 to: 1491

6 AsnArgAsnProProThrSerThrValArgAspAsnSerProPro 22
      ::::::::::::::::::::
49 AATGAGAACAGATGCCAGCGGTCAAGATGCGGGGAGCCCGAGGCC 98
      ::::::::::::::::::::
22 OGluProSerProGluSerLeuArgHisValSerArgMetIleAsnSera 39
      ::::::::::::::::::::
99 TGCCAGCTCCCAAGTGC..... 117
      ::::::::::::::::::::
39 snHISLTrpThrSerProSerArgThrIleTrpSerAspArgPheIlePro 55
      ::::::::::::::::::::
118 .....TGCCCAAGCAAG.....CAGGAGACCGCTTCATCC 150
      ::::::::::::::::::::
56 SerArgSerAlaSerLysPheAlaLeu.....PheAspIleAsn..... 68
      ::::::::::::::::::::
151 TCCAGAGCGCGGAGCCACTGAGCGTGACTCCACAGATTAACGAGAA 200
      ::::::::::::::::::::
69 .....ThrProThrGluGlyArg.....AspAspSerSer..... 79
      ::::::::::::::::::::
201 TGAGAGTCTCCAGTCAGAACCGGAAAGCCAGCGCCCTCAGACA 250
      ::::::::::::::::::::
80 .....AlaTrpThrThrLeuLeuArgThrAlaLeuPhe 90
      ::::::::::::::::::::
251 ACGGCAAGAGAGCGCTGCTACTCTCTCTCTCAAGATGAGCTGCTG 300
      ::::::::::::::::::::
91 Gly.....ProAspValAlaGlyProValThrProGluLysThrAs 104
      ::::::::::::::::::::
301 GTGCTCCGATCGAGAGGTGAGAGCCGACAGATGAGAGCGGAGCT 350
      ::::::::::::::::::::
104 pSerProSerMetThrLeuProAsnArgAsnIlePheArgTrpLysThrG 121
      ::::::::::::::::::::
351 GCAGCCCTCC...ACGCTGAGAAAGAGGTGTTGACATATCCCTTA 397
      ::::::::::::::::::::
121 LurThrArgGlnSerMet.....HisSer 128
      ::::::::::::::::::::
398 GCACCAAGCGCTCCAGCGCCGATGAGCGCAAGATGCTGCTCCCTAC 447
      ::::::::::::::::::::
129 LeuSerProPheMetLaspAspAspPheValProGlyValAsnHisSerPr 145
      ::::::::::::::::::::
448 CTGCTCTCCGCTCAGCAACAGAGC.....CAGAACGCTCCGCGTCC 491
      ::::::::::::::::::::
145 OValLysAlaProArgLysValProArgSerProTrpLysValLeuAspA 162
      ::::::::::::::::::::
492 CCGGAAACCCACCCGCAAGATCTCCAAAGATCCCTTCAAGGTGAGAG 541
      ::::::::::::::::::::
162 LArProAlaLeuGlnAspAspPheTrpLeuAsnLeuValAspTrpSerSer 178
      ::::::::::::::::::::
542 CGCGCGAGCTGCGAGAGCACTTACTCTCAATGCTGAGACGTGCTGCC 591
      ::::::::::::::::::::
179 HisAsnValLeuAlaValGlyLeuGlyAsnCysValTrpLeuTrpAsnAl 195
      ::::::::::::::::::::
592 CTCAAATGCTGCTAGCTGGGGCTAGGACACTGCTACTCTGAGAGTGC 641
      ::::::::::::::::::::
195 acySerSerLysValThrLysLeuGlyAspLeuGlyValAsp...AspC 211
      ::::::::::::::::::::
642 CTGTACACAGCAGTACCGCGCTGTGACCTTCACAGGAGGAGGACT 691
      ::::::::::::::::::::
211 yValLysSerValGlyTrpAlaGlnArgGlyThrHisLeuAlaValGly 227
      ::::::::::::::::::::
692 CAGTACACCTCCGCGGCTGCTGAGCGGGAGAACTGCTGCGGCTGGCC 741
      ::::::::::::::::::::
228 ThrAsnAsnGlyLysValGlnIleTrpAspAlaAlaArgCysLysLysI 244

```


[illegible]

AUTHORS
Strausberg, R.

TITLE
Direct Submission

JOURNAL
Submitted (27-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: gagbs+email.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (ILNL)
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalona@bcm.tmc.edu,
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.

FEATURES
source
1. 1861
/organism="Mus musculus"
/db_xref="locusID:56371"
/db_xref="taxon:10090"
/clone="MGC:11723 IMAGE:3967009"
/tissue_type="Mammary tumor. Metallothionien-TGF alpha model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI CGAP_Mam1"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"
224..1705
/codon_start=1
/product="Fzrl protein"
/protein_id="AAH0616.1"
/db_xref="gi:13879284"
/translation="MDQYERRLLRLIIIONIENTVPVCSEMRRTLPFANSPVSSPKRKHGDRIPPRAGANVMFNHRINENKSSPSONKADATSDNGKDIALSAIKNELDA
GIERVPPOTEDRLQLDPTPEHKGLFTYSLSKRSSPDGDNDVSYELSPYSNKSQKL
LRSPRRPKRIKSLIPFKVLDAEODDFYNLVWSSLNVLAVGLTCVTYMSACTSO
VTRLCDSVEGDSVTSGVSEGRGNLIVAVGHKGVOITWDAAGKLSMLECHTAHVAGA
LANWADLOSISRDMILLORDIRTPPLQSERLOGHQEVCGKMWSTPHOLASGNDL
NKLIAMHNSLSLPYOYTEHLAAVKALAMSPHOGILASGGGTADRCIRFWNTLTGDD
LOCIDGSOVCNCLAMSKHANVELYTHGHSNOQILVWYPSLTGYAKLIGHSYRVLYLA
MSPDEGAIVCAGEPTLRFNMVFESKTSTRKESVNLFFPIR"

CDS

BASE COUNT	409 a	560 c	562 g	330 t
ORIGIN				

alignment_scores:

Quality:	1286.00	Length:	495
Ratio:	3.448	Gaps:	14
Percent Similarity:	75.354	Percent Identity:	53.939

alignment_block:
US-09-701-572-2 x BC006616 ..

Align seg 1/1 to: BC006616 from: 1 to: 1861

```

6 AsnAtgAsnProProBProThrSerThrValAlaGAspAsnSerProPrp 22
|||||
|||::|||
||:::
272 AATTGACAAACAAGTGCCCTGTTCAGAGATGCCGAGAACCTGTGCAACC 321
|||||
|||
22 oGUUProSerProGluSerLeuArgHisValSerArgMetIleAnsSera 39
|||||
|||
322 AGCCAACATCCCCAGTGCT..... 340

```

```

39 snHsIytrHsrProSerArgThrIleTyrSerAspArgPheIlePro 55
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
341 .....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
56 SerArgSerAlaSerLysPheAlaLeu.....PheAspIleAsn..... 68
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
374 TCGCGGGCCGGGCGCACTGAGAGCTTCAACAGATCAATGAAAA 423
69 .....ThrProThrGluGlyArg.....AspAspSerSer..... 79
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
424 TTAGAAGTCCCCAGCAGAACCCGCAAGAGCCGACCTCTGGGACA 473
80 .....AlaTyrThrThrLeuLeuArgThrAlaLeuPhe 90
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
474 ATGGCAAGAGCGGCTGGCTACTCCGCACTGCTGAGATAGAGTGGT 523
91 Gly.....ProAspValAlaGlyProValThrProGluLysThrAs 104
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
524 GGTGGCGGCATCGAGAGGTTCCAGACCCACAGAGGAGCGCGGCT 573
104 pSerProSerMetThrLeuProAsnArgAsnIlePheArgTyrThrG 121
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
574 GAGGCGGTCC...ACACAGAGCAGACAGGGGCTCTTACGTAATCCCTCA 620
121 LuThrArgGlnSerMet.....HisSer 128
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
621 GGAGCAGAGCGCTCGAGTCCAGATGATGCAATGACGTGCCCACTACT 670
129 LeuSerProPheMetAspAspAspPheValProGlyValAsnHisSerPr 145
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
671 CTGTCCCGCTTACGCAACAAAGT.....CAGAGCTGTGGGCTCAC 714
145 ovalLysAlaProArgLysValProArgSerProThrLysValLeuAspA 162
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
715 ACCGAGGCCACACGCAAGATCTTAAAGATTCCCTCAAGGCTGGAGCG 764
162 IaProAlaLeuGlnAspAspPheTyrLeuAsnLeuValAspTyrSerSer 178
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
765 GCCCAGAGGCTTCAGAGCAGCTCTACCTCAACTTGTGGACTGGCTCC 814
179 HisAsnValLeuAlaValAlaGlyLysncysValTyrLeuTyrPasnAl 195
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
815 CTCAACAGGCTTCAGAGGCTGGGCGGCGGCTGCTGCTGAGAGTGC 864
195 acYSerSerLysValThrLysLysCysAspLeuGlyValAsp...AspC 211
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
865 ATGCACACAGCAGGTGACCCGGCTGTGACCTCTCTTAAAGGGGACT 914
211 ysaValCysSerValGlyTyrPalaGlnArgGlyThrHisLeuAlaValGly 227
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
915 CAGTACTTCCGTTGGTGTGCTGAGCGGGGAACTGTGTCGAGTAGT 964
228 ThrAsnAsnGlyLysValAlaGlnIleTyrPasnAlaValArgCysLysLysI 244
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
965 ACACACAAAGGCTTCTGCAATCTGGGAGCCTGCTGTGGGAAAGAACT 1014
244 earGserMetGluGlyHisArgLeuArgValAlaLeuAlaLratrPserS 261
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
1015 GTCCACTGTGGAGGGCCACACAGCAGAGTGGGGCGCTGGCTGGAGATG 1064
261 earSerLeuLeuSerSerGlyLysArgAspLysAsnIleTyrGlnArgAsp 277
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
1065 CTGACCACTGTGCACTGTGAGCCGTGACCCGATGCTTCAACAGCGGAT 1114
278 IleArgThr.....GlnGluAspPheValSerLysLeuSerGlyHI 291
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
1115 ATCCGACACACACCCCTGACGTACAG.....CGGGCGTGCAGGGCCA 1158
291 slySerSerGluValCysGlyLeuLysTyrPasnAlaValArgLysLeuA 308
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
1159 CCGGAGGAAAGTGTGTGGCTTAAAGTGTCCACAGACCAACCAAGCTGCTG 1208

```

```

308 1aSerGlyGlyAsnAspAsnLysLeuPheValTyrPasnGlnHisSerThr 324
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
1209 CCTGGGGGGGCAATGACACAAAGCTGCTGCTGGAACCACTAGTCTA 1258
325 GlnProValLeuLysTyrCysGlnHisThrAlaValLysAlaIleAl 341
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
1259 AGCCTGTGCAGCAGTATACGAGACACTGCGAGCGCTGAAAGCTATTGC 1308
341 atrPserProHisLeuHisGlyLeuLeuAlaSerGlyGlyTyrPala 358
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
1309 CTGGTCCCAACACACGATGACTGCTGCGCATCTGGTGGGCGACGCTG 1358
358 spArgCysIleArgPheTyrPasnThrThrTrasnSerHisLeuSerCys 374
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
1359 ACCGCTATCCGATTTCTGGAACACTCTGACAGTCAAGCAGCAGCTGC 1408
375 MetAspThrLysSerGlnValCysAsnLeuValTyrSerLysAsnValAs 391
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
1409 ATTGACACAGGCTTCACAAAGTGTGCAACCTGCGTGGTCCAAAGCCCA 1458
391 ngluLeuValSerThrHisGlyTyrSerGlnAsnGlnIleLeuValTyr 408
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
1459 TTAGCTGTGAGACACATGGCTACTCAGAAACAGATCTGTGTGG 1508
408 rglTyrProThrMetSerLysLeuAlaThrLeuThrGlnHisThrTyrArg 424
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
1509 AGTACCGCTCCCTTACGACAGTGGCCAAAGCTCACTGGCCACTATTCGT 1558
425 ValLeuTyrLeuAlaIleSerProAspGlyGlnThrIleValThrAla 441
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
1559 GTCTCTTACCTGGCCAGATGCTCCCTGATGGGAGCCATGATGCACCGAC 1608
441 aglyAspGluThrLeuArgPheTyrPasnValPheProSerProLysSerG 458
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
1609 TGGAGATGACACCTGAGGTTCTGGAATGTCTTC.....AGCAAAACAC 1652
458 lnaSnThrGlnSerGluIleGlyAlaLeuSerLeu 469
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCAATCCCC 373
1653 GCTCTACAAAGAAATGCTGTGTGTGCTCAACTC 1687

seq_name: gb_ro:AF083809
seq_documentation_block: 2258 bp mRNA linear ROD 01-SEP-1999
LOCUS AF083809
DEFINITION Mus musculus fizzy-related protein (Fyr) mRNA, complete cds.
ACCESSION AF083809
VERSION AF083809.1 GI:5813824
KEYWORDS
SOURCE
ORGANISM
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 2258)
JOURNAL Characterization of mouse fizzy-related protein
AUTHORS Jin,D.-Y. and Jeang,K.-T.
REFERENCE 2 (bases 1 to 2258)
JOURNAL Unpublished
AUTHORS Jin,D.-Y.
TITLE Direct Submission
JOURNAL Submitted (12-AUG-1998) LMM, NIAID/NIH, 9000 Rockville Pike,
Bethesda, MD 20892-0460, USA
FEATURES
source
location/Qualifiers
1..2258
/organism="Mus musculus"
/db_xref="taxon:10090"
1..2258
/gene="Fyr"
143..1624
/gene="Fyr"
/function="cell cycle regulator"
/codon_start=1
/product="fizzy-related protein"
/protein_id="AAD52029.1"

```


seq documentation block:

LOCUS XLRNAFRP 2649 bp mRNA linear VRT 08-SEP-1997
 DEFINITION Xenopus laevis mRNA for Fizzy-related protein.
 ACCESSION Y14163
 VERSION Y14163.1 GI:2326942
 KEYWORDS fizzy-related protein; far gene.
 SOURCE African clawed frog.
 ORGANISM Xenopus laevis

REFERENCE

AUTHORS Sigrist, S.J. and Lehner, C.F.
 TITLE 1 (bases 1 to 2649)
 JOURNAL 2 (bases 1 to 2649)
 MEDLINE 97433078
 REFERENCE Cell 90 (4), 671-681 (1997)
 AUTHORS Lehner, C.F.
 TITLE Direct Submission
 JOURNAL Submitted (04-JUL-1997) C.F. Lehner, Department of Genetics,
 University of Bayreuth, 95440 Bayreuth, FRG

FEATURES

source 1..2649
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /tissue_type="ovary"
 5'UTR 10..99
 gene 100..1581
 /gene="fzr"
 CDS 100..1581
 /gene="fzr"
 /codon_start=1
 /product="fizzy-related protein"
 /protein_id="CAA74576.1"
 /db_xref="GI:2326943"
 /translation="MDQYERLLRQINLQNTIPCASEMRRLTPSPMSPSKH
 GDRFISRAQAMSTINFGRIINERKSPQNRKAKATADSGDLAYSALKNELIGA
 GIEKVDPTEDERLQSTPEKSLPTYSLSKRAPDQCNVSPISLSPVNSKOKL
 LNSPKPRTRISKIPKVIDAPLQDFYLNLDVSLNVLVSGKISTLEGHTARVGA
 VTRLCDLSEGDVSTVSWSEKGNLAVGTHKGVAVDASAGKISTLEGHTARVGA
 LAMNADQLSGSRDMLQIDIRTPVQSEKRIQGRQVCGIKSTDHQLASGND
 NKLIVNHSLSPVQOYTEHLAAVKAIAKSPHOGHLAGSGTADRCIFPMWTLIGOP
 LQCIDGSOVCNLAISKHANEIYSTHGYSONOILWKYPSLQVAKLGHYSRVLYLA
 MSPDGRALVYAGADEFLRFNVPNSKRTKESVYANLFTFRIR"
 3'UTR 1582..2649

BASE COUNT 758 a 541 c 598 g 752 t
 ORIGIN

alignment_scores:

Quality: 1285.00 Length: 495
 Ratio: 3.427 Gaps: 14
 Percent Similarity: 75.758 Percent Identity: 53.737

alignment_block:

US-09-701-572-2 x XLRNAFRP ..

Align seg 1/1 to: XLRNAFRP from: 1 to: 2649

6 AsnATGAsnProProthTherTherValAlArGAspAsnSerProProPr 22
 |||:||||| ||| :|||||
 148 AATGAGAACACCATATCTTGTCATCAGAAATGAGAGACATTAAACCC 197
 22 oGluProSerProGluSerLeuAlGHisValSerArgMetIleAsnSera 39
 | ||||| |||
 198 AACCAACTCTCCCATGCT 216
 39 snHisTyrTherSerProSerArgThrlIleTyrSerAspArgPheIlePro 55
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 217TCACCAAGTAAA.....CATGAGACAGGTTTATACCA 249

56 SerArgSerAlaSerLysPheAlaLeu.....PheAspIleAsn..... 68
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 250 TCCGGGCGAGAGCGAATTGGACATTACTTCCACAGAAATTAATACAGA 299
 69ThrProThrGluGlyArgAsp.....Asps 77
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 300 TGAATAATCCCGACATCAGACAGAAAGCCAGATGCCAGCGCAGACA 349
 77 eSerSer.....AlaTyrThrThrLeuLeuArgThrAlaLeuPhe 90
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 350 GTGGCAAGATGCCCTAGCTTATTCAGCTCTGTTAAGATAGATGAGCTGTG 399
 91 Gly.....ProAspValAlaGlyProValThrProGluLysThrAs 104
 ||| ||| ||| :|||:|||||
 400 GGAGCTGGTATTGAGAAAGTTCAAGATCCACAGACGTAAGATAGACGACT 449
 104 pSerProSerMetThrLeuProAsnArgAsnIlePheArgTyrLysThrG 121
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 450 ACAACCATCA...ACGCCAGAAAAAAGAGCCTGTAGCTACTCTCTAA 496
 121 IuThrArgGln.....SerMetHisSer 128
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 497 GTTCTAAGCGTGCAGCCAGATGATGAATGAGCTGCTCTTATTTCT 546
 129 LeuSerProPheMetAspAspAspPheValProGlyValAlaHisSerPr 145
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 547 TTGTCACTCTCAGTAAATAAAC.....CAGACCTATTGGCGGTCTC 590
 145 oValLysAlaProArgLysValProArgSerProTyrLysValLeuAspA 162
 ||| ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
 591 ACGAAAGCCCTACAGAGAGATTTGGAAGATCCCTTTAAAGTCTGAGAG 640
 162 IabProAlaLeuGlnAspAspPheTyrLeuAsnLeuValAspTrpSerSer 178
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 641 CACCTGATTACAAGATGACTTTACTCTACCTACCTAGTGGATGGTATCC 690
 691 CTCATGTTTAAAGTGGGCTTGGAACCTGTGTTACTTATGAGCTGC 740
 195 acySerSerLysValThrLysLeuGlyAspLeuGlyValAsp...AspC 211
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 741 ATGTACTAGCAGAGTACTAGTATGATGATGATGATGATGATGATGATG 790
 211 yValLysSerValGlyTrpAlaGlnArgGlyThrHisLeuAlaValGly 227
 ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 791 CTGTACTTACTAGTTGGTGTGTCAGACCGGGAACCTGTAGCTGGGG 840
 228 ThrAsnAsnGlyLysValGlnIleTrpAspAlaAlaArgCysLysLysI 244
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 841 ACACACAAAGCTTTGTTGATGATTTGGATGATGATGATGATGATGATG 890
 244 eArgSerMetGluGlyHisArgLeuArgValGlyAlaLeuAlaTrpSerS 261
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 891 GTCCACTCTGAGGAGACACACACCCGCTGGGAGGAGCTTGGCTGGAAG 940
 261 eSerLeuLeuSerSerGlyLysArgAspLysAsnIleTyrGlnArgAsp 277
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 941 CAGACCAAGCTTCTCTCGGAGTAGGAGACAGATGATCTTACAAAGAGAT 990
 278 IleArgThr.....GlnGluAspPheValSerLysLeuSerGlyH 291
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 991 ATTGCGACACACCTGTTTCAGTCAGG.....AGAAAGCTCGAAGGCCA 1034
 291 sLysSerGluValCysGlyLeuLysTrpSerTyrAspAsnArgGluLeuA 308
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1035 CAGGCAAGAGTGTGGCTCAAGTGTGTCACAGACACACACCTCTGG 1084
 308 IAspGlyLysAspAsnLysLeuPheValTrpAsnGlnHisSerThr 324
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1085 CTTCAGAGGAGATGACAAACAGCTCTGTGGTTGGAACATTCTACCTCG 1134
 325 GlnProValLeuLysTyrCysGlnHisThrAlaAlaValLysAlaIleAl 341

```
1135 AGTCCGTGTCACAGTATATCTGAACATCTACCTGCAGTGAAGGCAATTGC 1184
341 aTrrSerProHISLeuHISGlyLeuAlaSerGlyGlyThrAla 358
1185 CTGGTCCCTCATCAACATGGCTTTGGCTTCAGCGGGGCACTGCTG 1234
358 sPArgCysIleArgPheTrpAsnThrThrAsnSerHISLeuSerCys 374
1235 ACCGCTGCATCCGCTTCTGGAACACACTGACAGCCAGCCCTTTACAGTGT 1284
375 MetAspThrGlySerGlnValCysAsnLeuValTrpSerLysAsnValAs 391
1285 ATTGATTAAGTGGCTCTCAAGCTGTAACTGGCCCTGCTAAGCATGCAGAA 1334
391 nGluLeuValSerThrHISGlyTyrSerGlnAsnGlnIleIleValTrp 408
1335 TGAACGTGTAAGCACACATGGCTACTCTCAAAATCAGATATTAGTATGGA 1384
408 rGlyTrpProThrMetSerLysLeuAlaThrLeuThrGlyHISThrTyrArg 424
1385 AATATCCATGCGCTTACTCAGGTGGCCAGTGTGACTGGACATTCTACCGT 1434
425 ValLeuTyrIleuAlaIleSerProAspGlyGlnThrIleValThrGlyAl 441
1435 GTTCTCTACCTGGCTATGTACACAGATGGGAGGCCATTGTTACAGGAGC 1484
441 aGlyAspGluThrLeuArgPheTrpAsnValPheProSerProLysSerG 458
1485 AGGTGATGAACGCTGAGATCTCTGGAATGTTTTT....AGCAAAACAC 1528
458 InAsnThrGluSerGluIleGlyAlaLeuSerLeu 469
1529 GCTCCACAAGAGAGTGTGTCTGGTTCYAAACCTG 1563
```



```

325 GlnProValLeuLysTyrCysGlnHisThrAlaAlaValLysAlaIleAl 341
      :::::::::: ::::: ::::: ::::: ::::: :::::
1297 ACCCCCGTCGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1346
      :::::::::: ::::: ::::: ::::: ::::: :::::
341 arPserProHISLeuHISGLYLeuLeuAlaSerGlyGlyThrAla 358
      :::::::::: ::::: ::::: ::::: ::::: :::::
1347 CTGGTCCCACTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1396
      :::::::::: ::::: ::::: ::::: ::::: :::::
358 sPaRqCysIleArpPheTrpAsnThrThrAsnSerHisLeuSerCys 374
      :::::::::: ::::: ::::: ::::: ::::: :::::
1397 ACCCGTGTATCCGCTTGTGACACAGCAGCAGCAGCAGCAGCAGCAGT 1446
      :::::::::: ::::: ::::: ::::: ::::: :::::
375 MetAspThrGlySerGlnValCysAsnLeuValTrpSerLysAsnValas 391
      :::::::::: ::::: ::::: ::::: ::::: :::::
1447 ATCGACACGGCGCTCCCAAGTGTGCAATGCGCTGTGCTCCACAGCCAA 1496
      :::::::::: ::::: ::::: ::::: ::::: :::::
391 nGluLeuValSerThrHisGlyTyrSerGlnAsnGlnIleIleValTrpA 408
      :::::::::: ::::: ::::: ::::: ::::: :::::
1497 CGAGCTGGTGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGT 1546
      :::::::::: ::::: ::::: ::::: ::::: :::::
408 rGTrProThrMetSerLysLeuAlaThrLeuThrGlyHisThrTyrArg 424
      :::::::::: ::::: ::::: ::::: ::::: :::::
1547 AGTACCCCTCCCTGACCCAGGTGCGCAAGCTGACGGGCGACCTCAACGC 1596
      :::::::::: ::::: ::::: ::::: ::::: :::::
425 ValLeuTyrLeuAlaIleSerProAspGlyGlnThrIleValThrGlyAl 441
      :::::::::: ::::: ::::: ::::: ::::: :::::
1597 GTGCTGTACTGTGCAATGTCCTGATGGGAGCCATGCTGACTGTGTC 1646
      :::::::::: ::::: ::::: ::::: ::::: :::::
441 aGlyAspGluThrLeuArgPheTrpAsnValPheProSerProLysSerg 458
      :::::::::: ::::: ::::: ::::: ::::: :::::
1647 TGGAGACGAGACCTGAGGTCTGTGACGCTTTAGCAAAACCCCTTGA 1696
      :::::::::: ::::: ::::: ::::: ::::: :::::
458 InsanThrGluSerGlu...IleGlyAlaLeuSerLeu 469
      :::::::::: ::::: ::::: ::::: ::::: :::::
1697 CAAGGTAAAGTGGAGTGTGTGTGCTGCTCAACCTC 1734
      :::::::::: ::::: ::::: ::::: ::::: :::::
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL05493
seq_documentation_block:
ID ABL05493 standard; cDNA; 2645 BP.
XX
AC ABL05493;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 10961.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PE 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW.
XX
DR WPI; 2001-656860/75.
XX
DR P-PSDB; ABB61390.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX

```

```

PS Claim 1: SEQ ID NO 10961; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB87737-AB872072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2645 BP; 766 A; 630 C; 619 G; 630 T; 0 other:

alignment_scores:
      Quality: 1276.00      Length: 474
      Ratio: 3.535      Gaps: 11
Percent Similarity: 76.160      Percent Identity: 54.852

alignment_block:
US-09-701-572-2 x ABL05493 ..

Align seg 1/1 to: ABL05493 from: 1 to: 2645

22 ProGluProSerProGluSerLeuArgHisValSerArgMetIleAsn 38
      :::::::::: ::::: ::::: ::::: ::::: :::::
507 CCCGAGTACAGAGAGCCAGCTGGAAGCACTACAGCTGCTGGCCAGAA 556
      :::::::::: ::::: ::::: ::::: ::::: :::::
38 r.....AsnHisTyrThrSerProSerArgThrIleTyrSerAspArg 53
      :::::::::: ::::: ::::: ::::: ::::: :::::
557 TCTGTTCACAACTTCGATGCTGTCACACTAGCCCACTCTCTGACCGCT 606
      :::::::::: ::::: ::::: ::::: ::::: :::::
53 helieProSerArgSer.....AlaSerLysPheAlaLeuPhe 65
      :::::::::: ::::: ::::: ::::: ::::: :::::
607 TCATACCCCTGCGAGCGCTACACAACTGCGACAGAACTTTGGC..... 650
      :::::::::: ::::: ::::: ::::: ::::: :::::
66 AspIleAsn.....ThrProThrGluGlyArg 74
      :::::::::: ::::: ::::: ::::: ::::: :::::
651 TCATATCAACAACTCCAAATGACACTGCCCGCAGACAGTAGAAGCAGCG 700
      :::::::::: ::::: ::::: ::::: ::::: :::::
74 gAsp.....AspSerSerAlaTyrThrThrLeuLeuA 86
      :::::::::: ::::: ::::: ::::: ::::: :::::
701 GGACTGGGGGGAAGCGCAGCATAGTCTGCGCTACTCTGCTCTGCTGCA 750
      :::::::::: ::::: ::::: ::::: ::::: :::::
86 rGThrAlaLeuPheGlyProAspVal.....AlaGlyPro 97
      :::::::::: ::::: ::::: ::::: ::::: :::::
751 AGAAGCAGCTCTCTGCGCAATCGACAGCAGTGAAGACCGCGCGGAG 800
      :::::::::: ::::: ::::: ::::: ::::: :::::
98 ValThrProGluLysThrAspSerProSerMetThrLeuProAsnArgas 114
      :::::::::: ::::: ::::: ::::: ::::: :::::
801 GAGCGGAATGAGATGCTTACACCGCGCC.....GCAAGCGGAG 841
      :::::::::: ::::: ::::: ::::: ::::: :::::
114 nIlePheArgTyrLysThrGlnThrArgGlnSerMet..... 126
      :::::::::: ::::: ::::: ::::: ::::: :::::
842 TCTCTTCAAGTACACATCCACCACCAAGCAGACTACATAGCGAGTGTG 891
      :::::::::: ::::: ::::: ::::: ::::: :::::
127 ..HisSerLeuSerProPheMetAspAspPheValProGlyValAsn 142
      :::::::::: ::::: ::::: ::::: ::::: :::::
892 CGTACTCTGTGTACCCGTCAGCGCCCAAAAGT.....CAGAAGCTGTG 935
      :::::::::: ::::: ::::: ::::: ::::: :::::
143 HisSerProValLysAlaProArgLysValProArgSerProTyrLysVa 159
      :::::::::: ::::: ::::: ::::: ::::: :::::
936 CGATCGCCCGCCAGAGCTACGCGCAAAATCTCTGCGATTCCTTCAAGT 985
      :::::::::: ::::: ::::: ::::: ::::: :::::
159 IleuAspAlaProAlaLeuGlnAspAspPheTyrLeuAsnLeuValAsp 176
      :::::::::: ::::: ::::: ::::: ::::: :::::
986 GCTAGACGGCGCGAGTTGACAGCAGCTTCTATCTCAACCTGTGTCAGT 1035
      :::::::::: ::::: ::::: ::::: ::::: :::::
176 rpSerSerHisAsnValLeuAlaValGlyLeuGlyAsnValTyrLeu 192
      :::::::::: ::::: ::::: ::::: ::::: :::::

```



```

1423 AATGAGAACAGATGCCACGCGTCACAGAGATCGCGGAGACCTTGACGCC 1472
22  oGIProSerProGIuSerLeuArgHisValSerArgMetIleAsnSera 39
1473 TGGCAGCTCCCGCAGTGTCC..... 1491
39 snHisTyThrSerProSerArgThrIleTySerAspArgPheIlePro 55
1492 .....TCGCCAGCAAG.....CAGGAGACCGCTTCATCCCC 1524
56 SerArgSerAlaSerLysPheAlaLeu.....PheAspIleAsn..... 68
1525 TCCGAGAGCGGAGCCCACTGGACCGTGAACCTTCACAGGATTAAACAGAA 1574
69 .....ThrProThrGIuGIYArg.....AspAspSerSer..... 79
1575 TGGAGAGTCTCCAGTCAGAACCGGAAACCCAGGACGCCACCTCAGACA 1624
80 .....AlaTyThrThrLeuLeuArgThrAlaLeuPhe 90
1625 ACGGCAAAAGACGGCTGGCCTACTGCTGCCGTGTCAGAAATGAGCTGTG 1674
91 GIY.....ProAspValAlaGIYProValThrProGIuLysThrAs 104
1675 GGTGCCGCGATCGAGAGGTGCAAGACCCGACACTGAGAGCGGACGCGCT 1724
104 pSerProSerMetThrLeuProAsnArgAsnIlePheArgTYLysThrG 121
1725 GCAGCCCTCC...ACGCCGAGAAAGAGGTGTGTTCACGTATTCCTTA 1771
121 LuThrArgIuSerMet.....HisSer 128
1772 GCACCAAGGCGTCCAGCCCGGATGACGCGCAAGATGTCTCCCTACCTCC 1821
129 LeuSerProPheMetAspAspAspPheValProGIYValAsnHisSerPr 145
1822 CTGTCTCCCGTCAGCAACAAGAC.....CAGAACTGTCTCCGCTGCC 1865
145 oValLysAlaProArgLysValProArgSerProTYLysValLeuAspA 162
1866 CCGGAAACCCACCGGAGATCTCCCAAGATCCCTTCAAGGTGCGGAGAC 1915
162 lAProAlaLeuGIuAspAspPheTYLLeuAsnLeuValAspTrpSerSer 178
1916 CCGCCGAGTGCAGAGCGACTTCTACCTCAATCTGTGACGTGCTGCTCC 1965
179 HisAsnValLeuAlaValGIYLeuGIYAsnCysValTYLLeuTrpAsnAl 195
1966 CTCATGTCTCAGCGTGGGCTAAGCACCTCGTGTACTGTGGAGTGC 2015
195 aCysSerSer..... 198
2016 CTGTACACAGCCAGAAAGGAGAGCGGATGGAGAGCGGATGGGTAGACA 2065
198 ..... 198
2066 TGGGTGAGCGGATGGAGAGCGCATGGAGAGCGGATGGAGAGCGGATG 2115
198 ..... 198
2116 GGAGAGCCGATGGAGAGCGCATGGATGAGCGCATGGAGAGCGCATGGA 2165
199 .....LysV 200
2166 CTGGGGCTTCGAGACGCGCAAGCTCAGCTGACCTTGCCTCCAGG 2215
200 aIThrlLysLeuCysAspLeuGIYValAsp...AspCysValCysSerVal 215
2216 TGAACGCGGTCTGTGACTCTCAGTGAAGGGGACTCAGTGCCTCCGCTG 2265
216 GIYTrpAlaGIuArgLY..... 221
2266 GCGTGTCTGAGCGGGGTGAGGTGGGCTTGTAGTTCATTTAAGATTATGA 2315

```

```

222 .....ThrHisLeu..... 224
2316 TGAAAACATCTCAGCAGCAGCTTTGAACATATATATGAGGAGGGGAAAC 2365
225 .....AlaValGIYThrAsnAsnGIYLysValGIuIleTrpAspAla 239
2366 TGTGTGCGGTGGGCACACAAAGGCTTCGTGAGATCTGGAGAGCGACC 2415
240 ArgCysLysLysIleArgSerMetGIuGIYHisArgLeuArgValGIYAl 256
2416 GCAGGGAAGAGCTGTCATGTTGTGAGGGCCACACGCGACGCTCGGGGC 2465
256 aLeuAlaTrpSerSerSerLeuLeuSerSerGIYArgAspLysAsnI 273
2466 GCTGGGCTGGAAATGTGACAGCTGTGTGTGCGGAGCCGCGACGCTGA 2515
273 lETyrgIuArgAspIleArgThr.....GIuIuAspPheValSer 286
2516 TCTGAGAGGAGGACATCCGACCCCGCAGTGCAGTGCAG.....CGG 2559
287 LysLeuSerGIYHisLysSerGIYValCysGIYLeuLysTrpSerTYrAs 303
2560 CGGCTGCAGGGGCACCGCAGAGAGTGTGGGCTCAAGTGTCCACAGA 2609
303 pAsnArgGIuLeuAlaSerGIYGIYAsnAspAsnLysLeuPheValTrpA 320
2610 CCACAGCTCTCGCTCGGGGGGCAACGACACACAGCTGTGTGTGGA 2659
320 snGIuHisSerThrGIuProValLeuLysTYrCysGIuHisThrAla 336
2660 ATCACTGAGCTGAGCCCGCTGCAGCAGTAAACAGGACCTGGCGGCC 2709
337 ValLysAlaIleAlaTrpSerProHisLeuHisGIYLeuLeuAlaSerGI 353
2710 GTGAGGCGCATGCGCTGTGATCCCAATCAGACAGCGGCTGTGCGCTGGG 2759
353 YGIYGIYThrAlaAspArgCysIleArgPheTrpAsnThrThrThrAsn 370
2760 GGGGCGCAGAGCTGACCGCTGTATCGCTTGTGAAACAGCTGACAGGAC 2809
370 eHisLeuSerCysMetAspThrGIYSerGIYValCysAsnLeuValTrp 386
2810 AACCACTGCAGGTGTATGACAGCGGCTCCCAAGTGTGCAATCTGGCGTGG 2859
387 SerLysAsnValAsnGIuLeuValSerThrHisGIYTySerGIuAsnGI 403
2860 TCCAAACAGCAGCAAGAGCTGTGTAGCAGCCAGCGCTACTACAGAACCA 2909
403 nIleIleValTrpArgTYrProThrMetSerLysLeuAlaThrLeuThrG 420
2910 GATCCTGTGTGGAATGACCCCTCCGACCCAGGTGAGGACAGCTGACCG 2959
420 LysHisThrTYrArgValLeuTYLLeuAlaIleSerProAspGIYGIuThr 436
2960 GGCACCTCTACCGGCTGTGTACTGTGCAATGTGCCCTGATGGGAGGCC 3009
437 lIleValThrGIYAlaGIYAspGIuThrLeuArgPheTrpAsnValPhePr 453
3010 ATGCTCAGCTGTGTGAGAGCAGAGACCTGAGGTCTGGAACCTCTTTAG 3059
453 oSerProLysSerGIuAsnThrGIuSerGIuIleGIYAlaLeuSerLeuG 470
3060 CAATAACCGCTTCGAAAGAGGACAGCTTCCCTCCGAGG...CTGCGGCTCA 3106
470 LYArgThrThrIle 474
3107 GCGCATCTGCCATC 3120
seq_name: /sidsl/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL05492
seq_documentation_block:
ID ABL05492 standard; cDNA; 14283 bp.

```

```

XX AC ABL05492;
XX XX
XX 26-MAR-2002 (first entry)
XX XX
DE Drosophila melanogaster expressed polynucleotide SEQ.ID NO 10958.
XX XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX XX
XX Drosophila melanogaster.
XX XX
XX WO200171042-A2.
XX XX
XX 27-SEP-2001.
XX XX
XX 23-MAR-2001; 2001WO-US09231.
XX XX
XX 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX XX
XX (PEKE ) PE CORP NY.
XX PA
XX Venter JC, Adams M, Li PWD, Myers EW;
XX PI
XX WPI; 2001-656860/75.
XX DR
XX P-PSDB; ABB61389.
XX XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX PS
XX Claim 1; SEQ ID NO 10958; 21pp + Sequence Listing; English.
XX XX
XX The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
XX CC sequences (AB101840-AB16175) and the encoded proteins
XX CC (AB57737-AB72072).
XX CC
XX The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX CC
XX Sequence 14283 BP; 4076 A; 2905 C; 2990 G; 4312 T; 0 other;
XX SQ

```

alignment_scores:

```

Quality: 1156.50      Length: 380
Ratio: 3.894          Gaps: 5
Percent Similarity: 78.158      Percent Identity: 59.211

```

alignment_block:

```

US-09-701-572-2 x ABL05492 ..

```

```

Align seg 1/1 to: ABL05492 from: 1 to: 14283

```

```

115 11lephargtyrlysthrghutrrarglnsermethisserieser 131
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
11413 GTCCTTTCGTTGACGACTACAGTGGCGAGTGCCTACTCGTTCACCC 11462
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
131 opmetasppaspphevalproglyvalasnhisserieserprovalysa 148
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
11463 GTCGAGCGCCAAAGT.....CAGAAGCTGTTCGATCCGCCGCGCAAG 11506
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
148 laProAlglyValProArgSerProTyrLysValLeuAspAlaProAla 164
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
11507 CTACGCGCAAAATCTCTCGCATTCCTCAAGGTGCTAGACGCGCGCGAG 11556
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
165 LeuGlnAspSppheTyrLeuAsnLeuValAspTyrSerSerHisAsnVa 181
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

```

11557 TTGCAGACGACTTCTATCTGAACCTGGTCGAGTGTCTGCAGAAAGT 11606
181 1leuAlaValGlyLeuGlyAsnCysValTyrLeuTrrpAsnAlaCysSers 198
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
11607 ACTGCTGTAGGCTGGCTGGCAGCTGTGTATCTGTGTGAGCGCGTGCACCA 11656
198 er..... 198
11657 GTACAGTGAATCTTGTATCATAGGCAATATGCTTCACCTTGTGTTCATA 11706
199 .....Lys..ValThrLysLeu 203
11707 AACTAATCAAAATGATTTTCCGATCGATATGATGATGATGATGATGATG 11756
204 CysAspLeuGlyValAsp...AspCysValCysSerValGlyTrrpAlaG1 219
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
11757 TGTGATCTCATGTCGGATCGGATCGGAATAGCGTGAACCTGCTGTGGAACGA 11806
219 nArgGlyThrHisLeuAlaValGlyThrAsnAsnGlyLysValGlnIleT 236
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
11807 GCGTGGCAACACCGCGCGCGGCGACACATACGCGTACGTCGACCGCTCT 11856
236 rPAspAlaAlaArgCysLysLysIleArgSerMetGluGlyHisArgLeu 252
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
11857 GGGATGTGGCGCCCAATAGACAGATCAACAACCTGAATGCGCATTCGCGCG 11906
253 ArgValGlyAlaLeuAlaTrrpSerSerLeuLeuSerSerGlyGlyArg 269
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
11907 CGTGTGGCGCGCTTGGCATGAGACAGTCAATCCGTGTGACGCGGTGCGG 11956
269 gAspLysAsnIleTyrGlnArgAspIleArgThr...GlnGluAspPheV 285
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
11957 AGACCGTTGATCATACACAGCGGATACAGAAACGCGCACTGCAATCGG 12006
285 alserLysLeuSerGlyHisLysSerGluValCysGlyLeuLysTrrpSer 301
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
12007 AGCGGAGATGGCGCGACATCGCAGAGGTGTGCGGCTGAATATGCTCA 12056
302 TyrAspAsnArgGlyLeuAlaSerGlyLysAsnAspAsnLysLeuPheVa 318
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
12057 CCGATATATCAATACTTGGCCGATGGCGGCAACGATATCGGTGTATGT 12106
318 1TrrpAsnGlnHisSerThrGlnProValLeuLysTyrGlySerGlnHisTrr 335
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
12107 GTGGATCAGCATTCCTCGAATCCCGTAAATCAATACAGGAGCATATAG 12156
335 1aAlaValLysAlaIleAlaTrrpSerProHisLeuHisGlyLeuLeuAla 351
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
12157 CGGCTGTAAAGCGCATCCGCGTGTGCGCGCATACACAGCATCTGTGCC 12206
352 SerGlyGlyGlyThrAlaAspArgCysIleArgPheTrrpAsnThrThr 368
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
12207 AGCGCGGTGGGAACGGCGGATAGTGTATCCGTTCTGGAATACGCTGAC 12256
368 rAsnSerHisLeuSerCysMetAspThrGlySerGlnValCysAsnLeuV 385
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
12257 GGGCCAGCCCATGTCAGTGGGACACGCGCTCGCAGGTTTGCAATCGG 12306
385 alTrrpLysAsnValAsnGlnLeuValSerThrHisGlyLysSerGln 401
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
12307 CCGTGTCAAGACATCTCTGGAGCTGTGCTCCACGACGCTACTCTCGAG 12356
402 AsnGlnIleIleValTrrpArgTyrProThrMetSerLysLeuAlaThrLe 418
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
12357 AACGAGATACCTGTGTGAATATCCCTCCGACGCAAGTGGCCAACT 12406
418 uThrGlyHisThrTyrArgValLeuTyrLeuAlaIleSerProAspGly 435
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
12407 GACGGCCATTCGTGCTGTGCTCTATCTGCGCTGAGTCCGATGGTG 12456
435 1nThrIleValThrGlyAlaAspGlyLysThrLeuArgPheThrAsnVal 451
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
12457 AGCTATTGTTACGGCGCGCGGACAGACGCTGCGATTTTGGACACTA 12506

```

452 pheProSerProlYssSergInasThrGluSerGluIle 464
||| :::: ||||| :::::::::::
12507 TTCACGAAGGC GCGCAGTCAGAGAAGACAAGTCCGT 12545

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL18422

```
seq_documentation_block:
ID   ABL18422 standard; DNA; 3369 BP.
```

AC ABL18422;

DT 26-MAR-2002 (first entry)

DE *Drosophila melanogaster* genomic polynucleotide SEQ ID NO 6739.

kw Drosophila; developmental biology; cell signalling; insecticide
kw pharmaceutical; gene; ds.

05 *Drosophila melanogaster*.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231

PR 23-MAR-2000; 2000US-191637P.
 PR 11-THI-2000; 2000US-061415O

XX
DA (DEKE) DE COBB NY

XX	
PI	Venter TC Adams M

XX
WB1: 3001-555950/75
DP

XX
XX
PT

PT genes from *Drosophila* and for elucidating cell signalling and cell-cell interactions -

XX
PS
C) a) m 1: SEO TD NO 6739: 21m + \$ecurcccc Riet+jnr: End]sch

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from *Drosophila*. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the design of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB161616-AB130511), expressed DNA sequences (AB101840-AB161615) and the encoded proteins (AB857737-AB872072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pct_sequences](http://www.wipo.int/pub/published_pct_sequences).

Sequence 3369 BP; 802 A; 899 C; 924 G; 744 T; 0 other;

alignment_scores:

Quality:	997.00	Length:	461
Ratio:	3.125	Gaps:	10

alignment_block:

US-09-701-572-2 x ABL18422

Align seg 1/1 to: ABL18422 from: 1 to: 3369

52 Argphelleproserargseralaserlyspheialaleupheaspilleas 68
||||| :||| :||| :||| :|||
1083 CGCGCATTCGCCACAGAGCTGTGTAACCACTGGAACCCGGAATTGTCAA 1133
68 ntHrProThnGlucIyArgAspAspSerSereralatyrThrTrnLeuL 85
||||| :||| :||| :||| :|||

```

355 lYtHAlaAspRgCysIlleArpHeTtPaSnThrThrAsnSerHis 371
|||||
2011 GGTCCGCGGATCGTCTGCTTGGACGTCCTTGAACGCGGGAAGCTG 2060
372 LeuSerCysMetAspThrGlySerGlnValCysAsnLeuValTPrSerLy 388
|||||
2061 GTGAAGTGCATACACCGCGGCCCGCATGACAACTCGGCTGGGCGCAG 2110
388 sAsnValAsnGlnLeuValSerThrHisGlyTyrSerGlnAsnGlnIleI 405
|||||
2111 GGATTCGCCAGAGCTGGTATACACCGACGACGCGCAGCGACGAGTCA 2160
405 lValTTPArGlyTyrProThrMetSerLyLeuAlaThrLeuThrGlyHis 421
|||||
2161 TTGCGCTGGCGTATCCCTCGTGAAGCAGATGGCCAGGCTGCGGGGAC 2210
422 ThrTyrArgValLeuTyrLeuAlaIleSerProAspGlyGlnThrIleVa 438
|||||
2211 ACACAGCGGGTACTCCACTGTCAGTGAAGTCCGGATACGATCATATGT 2260
438 lThrGlyAlaGlyAspGluThrLeuArgPheTrpAsnValPhe.ProSer 454
|||||
2261 GACTGGCGGCGCAGATGAGACCGCTTCGCTTGTGACCGCTTTTACCAAC 2310
455 .....ProLySerGlnAsn..... 459
2311 AGAAATATTCAAAGACCAACTCCGCTCCGTTTGTAAAGCATA 2360
460 .ThrGluSerGluIleGlyAlaLeuSerLeu 469
|||||
2361 AGATGAGGACAAATATGCGACGATATCTTGG 2391
seq_name: /SIDSL/gcdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL05448
seq_documentation_block:
ID ABL05448 standard; cDNA; 5058 BP.
AC ABL05448;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 10826.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PE 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PMD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
DR P-PsDB; ABB61345.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 10826; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is

```

CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABBS737-ABBS7202).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SO Sequence 5058 BP; 1244 A; 1301 C; 1304 G; 1209 T; 0 other;

alignment_scores: Quality: 997.00 Length: 461
 Ratio: 3.125 Gaps: 10
 Percent Similarity: 69.197 Percent Identity: 44.685

alignment_block:
 US-09-701-572-2 x ABL05448 ..

Align seg 1/1 to: ABL05448 from: 1 to: 5058

```

52 ArgPheIleProSerArgSerAlaSerLySphenAlaLeuPheAspIleAs 68
|||||
2354 CGCTGATTCCTCCACACGAGCTGTACCCACTGGAAACCGAACTTTGTCAA 2403
68 nThrProThrGluGlyArgAspAspSerSerSerAlaTyrThrThrLeu 85
|||||
2404 TGCTCCGGAGACACTCCATGTGTACGAGAGCTGTGTACTCATGCTCT 2453
85 euArgThrAlaLeuPheGlyProAspValAlaGlyProValThrProGlu 101
|||||
2454 TGCGGCAAC..... 111
102 LysThrAspSerProSerMetThrLeuProAsnArgAsnIlePheArgTy 118
|||||
2465 TTCTCGGACGAGAACATCATCTCTTGAGAAC..... 2497
118 lYsThrGluThrArgGlnSerMetHisSerLeuSerProPheMetAspA 135
|||||
2498 ....TGCAGCAGACCGCAGAGTTCATGCC.....CAGACTG 2531
135 sPAspPheValProGlyValAsnHis..... 143
2532 ATGATTTTAAAGGATCGCATTAACCGTATCTTGTGTTTCCAGCGGTTCT 2581
144 ...SerProValLysAlaPro..... 149
2582 TCGTCGCCAGTACGACACACCATTTGCCAACATCAGCAGAGCCACAGC 2631
150 ..... 150
2632 GCTGCTCAGCATCTTCAAGAGACCCATGCGCGACTGCGCGCTCCCGT 2681
157 yrlYsValLeuAspAlaProAlaLeuGlnAspAspPheTyrLeuAsnLeu 173
|||||
2682 ACAGATGATCTGATGCGCGGAGCTGCAAGACGACTTCTACTGACACTT 2731
174 ValAspTyrSerSerHisAsnValLeuAlaValGlyLeuGlyAsnCysVa 190
|||||
2732 ATTGACTGCTGCTCGAATAAACCTCGCGCTGCTCGGCTGCTCGCT 2781
190 lYrLeuTrpAsnAlaCysSerSerLyValThrLysLeuCysAspLeu 206
|||||
2782 TTATCTCTGAGTGGCGTCAAGTGAAGTGAAGGACTATGCGACTTCA 2831
207 ..GlyValAspAspCysValCysSerValGlyTyrPalGlnArgGlyThr 222
|||||
2832 ACAATGAGATTAACCTGATTACAGCAGTATGCTGCAGCGTGAAGCGCG 2881
223 HisLeuAlaValGlyThrAsnAsnGlyLysValGlnIleTrpAspAlaAl 239
|||||

```

```

2882 CAGGTGGCCATTGGACACCCAGTCCGGGTACGTGACCATTTGGATGCCGA 2931
2931 aargcylslyslsleargsermetgluglhisargleargvalgla 256
2932 GAACCAAAACACATAAACCGACTGGAGAGACACTGGCGCGGTGTAACGG 2981
256 laleualatrpserserleuleuenserllyllyargasplyasn 272
2982 CGCTACTGTTGGCGCAACCGATGTGGCCAGTGTCCCGGACCCCTCA 3031
273 lletyglargaspilleargthrglgluasphavalserlys...le 268
3032 ATTTCGAGGGGATATCCGCAATCCGCCAGCACATTAACCCGCGCT 3081
288 userglyhslslysergluvalcysglyleuylstrpsertyraspa 305
3082 GCCGGACACAAAGCTAGAGGTGTGTGCTTACAGTGTGCCAGCAAC 3131
305 rgluleualaserglyglyasnaspasnllysleuphevaltrpsn 321
3132 GATACCTGGCCAGCGGTGGCAGGACACACCGCTGTGGTTGGACCC 3181
322 Hlaserthrglprovalleuylstrpsertyrasgluhsrthlala 338
3182 GATTGGCCGAGCCCATTTACCCCTTCGACGACACACAGCGGTGCTA 3231
338 salalialatrpserprohisleuhsiglyleuualaserglylg 355
3232 GCGCTGGCGGTGGTGGCGCATATAGTCGGTCTTGGCCATGGCGGG 3281
355 lythralaspargcysleargpmetrpasrthtrthrasnserhis 371
3282 GGTCCGCGATGCTGCTGCTGCTTTGGAAAGCTCTGACCGGAAGCTG 3331
372 leuSerCysMetAspThrGlySerInValCysAsnleuValtrpser 388
3332 GTGAAGTGCATCAACACCGCGCCGACATACACACACTGGCTGGCC 3381
388 sasnlvalasngluleuvalserthrglytyrsergluasnnglile 405
3382 GGATTCGCGAGAGCTGTGTACACACCGACGACCGCACCGCACGCA 3431
405 levaltrpartrpthrmetserlysleualatthleuThrglyhis 421
3432 TTGCTGGCGCTATCCCTCGCTGAAGCAGATGGCCAGGCTGCGGGAC 3481
422 ThrTrpArgValleuThrglyleualalieserProAspGlyInThrIle 438
3482 ACACAGCGGGTACTCCACCTGTCACTGATCGGATACGATCATATGT 3531
438 lThrglylaaglyaspgluThrglyleuargpmetrpasnlphe... 454
3532 GACTGGCGCGAGTGTGAGACGCTGCTGCGACCGCTTTTCAAAAC 3581
455 .....ProlySerGlnasn..... 459
3582 AGAAATATTCAAAGACCAACCTCCGCTTGTGTTAAAGCATTA 3631
460 .ThrglysergluilegylalaleuSerleu 469
3632 AGATGAGAGCAAAATGTGCGAGCTATCTTTG 3662
seq_name: /STDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL18423
seq_documentation_block:
ID ABL18423 standard; DNA; 1369 BP.
XX
AC ABL18423;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 6742.
XX

```

```

KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
OS Drosophila melanogaster.
PN WO200171042-A2.
PD 27-SEP-2001.
PE 23-MAR-2001; 2001WO-US09231.
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
WIPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 6742; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB161737-AB172072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1369 BP; 304 A; 394 C; 394 G; 277 T; 0 other:
XX
alignment_scores:
Quality: 995.50 Length: 441
Ratio: 3.211 Gaps: 9
Percent Similarity: 70.295 Percent Identity: 45.805
XX
alignment_block:
US-09-701-572-2 x ABL18423 ..
XX
Align seq 1/1 to: ABL18423 from: 1 to: 1369
XX
52 ArgPheIleProSerArgSerAlaSerLysPheAlaLeuPheAspIleAs 68
83 CGCTGCATTCCTCCACAGCAGCTGTGTAACCACTGGAACCGCACTTTGTCA 132
68 nThrProThrglgluGlyArgAspAspSerSerSerAlaYrThrTrpLeuL 85
133 TGTCTCGGGAGACTCCATGTGTGACAGAGCTGTGTGTCATGCTGCTCT 182
85 euArgThrAlaLeuPheGlyProAspValAlaGlyProValThrProGlu 101
183 TGGGCAAC.....GAG 193
102 LysThrAspSerProSerMetThrLeuProAsnArgAsnIlePheArgTyr 118
194 TTCCTGACGAGACATCACTTCCTTGGAGAAC..... 226
118 rlystrhglnThrArgGlnSerMetHisSerLeuSerProPheMetAspA 135
227 ....TTCGAGCAGACCGCACAGGTCATGCC.....CAGACTG 260
135 spAspPheValProGlyValAsnHis..... 143

```



```

261 ATGATTTTAGGATGCGCATACCGTATCTTGTGTTGCCAGCGTCTCT 310
144 ... SerProValLysAlaPro..... 149
311 TCGTGGCCAGTACGACACCATTCGCCAACATCAGCAGAGCAGCCAGC 360
150 ..... ArgLysValProArgSerProT 157
361 GGTGCTCAGCATCTTCAGAGACCCATGCGCGACTGCCAGCTGCCCTCT 410
157 ytlYsValLeuAspAlaProAlaLeuGlnAspPheTyrLeuAsnLeu 173
411 ACAAGTACTGATCGCGCGAGCTGCAAGAGACTTCTACTGTAACCTT 460
174 ValAspTyrSerSerHisAsnValLeuAlaValGlyLeuGlyAsnGly 190
461 ATGAGTGTGCTGCGAAACACCTGCGCGTGTGCTGCGTGTGCTGCT 510
190 lYrLeuTyrPasnAlaCysSerSerLysValThrLysLeuCysAspLeu 206
511 TTAATCTGTGAGTGGCGTCAAGTGTGAGTGTGAGTGTGAGTGTGAGT 560
207 .. GlyValAspAspCysValCysSerValGlyTyrPalaGlnArgLys 222
561 ACAATGAGATTAACCTGATTACAGCAGTTCAGTTCAGTTCAGTTCAG 610
223 HisLeuAlaValGlyThrAsnAsnGlyLysValGlnLeuTyrPalaAla 239
611 CAGGTGGCAATTTGGCACCAGTCCGGTACGTACCATTTGGGATGCCA 660
239 aArgCysLysLysLeaArgSerMetGlnGlyHisArgLeuArgValGly 256
661 GAACCAAAAGCAGATTAACCGATCGAGACAGCATCGCGCGTGTACCG 710
256 lAlaLeuAlaTyrSerSerSerLeuLeuSerSerGlyLysArgAspLys 272
711 CGTACTGTGTGTGCGGCAACCGATTCGCGCGTGTGCGCGTGTGCG 760
273 lLeuTyrGlnArgAspLeuArgThrGlnGlnLysPheValSerLys... 288
761 ATTTTTCGCGGAGATTCGCGCATTCGCGCGAGCAGCATATACCGCG 810
288 uSerGlyHisLysSerGlnValLysGlyLysLysLysLysLysLys 305
811 GCGCGGACACAGCTAGAGGTGTGCGCTACAGTGTGCGCGCAGCAATC 860
305 rglLLeuAlaSerGlyGlyAsnAspAsnLysLeuPheValTyrPasn 321
861 GATACTTGGCGCGGTGCGAGCGCAGACACCGCTGTGTTGTTGCGCG 910
322 HisSerThrGlnProValLeuLysTyrCysGlnHisThrAlaAlaVal 338
911 GATTGGCGGAGCGCATTTACGCTTCGACGACACACAGCGGTGTGCA 960
338 sAlaLeuAlaTyrSerProHisLeuHisGlyLeuLeuAlaSerGlyGly 355
961 GCGCGTGGTGTGCGCGCATATAGTCCGCTTGTGCGCGAGTGTGCG 1010
355 lYrThrAlaAspArgCysLysArgPheTyrPasnThrThrTrpAsnHis 371
1011 GGTGCGCGCATCGCTGTGCTGTGGAACGCTTCACCGGAGAGCTG 1060
372 LeuSerCysMetAspThrGlySerGlnValCysAsnLeuValTyrSer 388
1061 GTAGAGTGCATCAGACACCGCGCGCGCATGACAGCACTGCGCGCAG 1110.
388 sAsnValAsnGlnLeuValSerThrHisGlyTyrSerGlnAsnGlnLeu 405
1111 GGAATCCCGAGAGCTGTGATCCACCGCAGCAGCAGCAGCAGGTCA 1160
405 lValTyrArgTyrProThrMetSerLysLeuAlaThrLeuGlnHis 421
1161 TTGCTGGCGCTATCCCTCGCTGAAGCAGATGCGCAGGCTGTCCGCG 1210

```

```

422 ThrTyrArgValLeuTyrLeuAlaLeuSerProAspGlyGlnThrIle 438
1211 ACACAGCGGGGACTCCACCTGTGATGATCGGATTAACGATCAATAGT 1260
438 lThrGlyAlaGlyAspGlnThrLeuAlaGlyPheTyrPasnValPhe 454
1261 GACTGCGCGCGCAGATGAGAGCGTTCGCTGTGACCGCTTTTACCAAC 1310
455 ..... ProlYsSerGln 458
1311 AGAAATATATCCAAAGACCCAA 1331
seq_name: /sids1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL07411
seq_documentation_block:
ID ABL07411 standard; cDNA; 1706 BP.
XX
AC ABL07411;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 16715.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN NC0200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001MO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PMD, Myers EW;
XX
DR WPI: 2001-656860/75.
DR P-PSDB; ABB63308.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 16715; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB161676-AB130511), expressed DNA
CC sequences (AB101840-AB161675) and the encoded proteins
CC (ABBS7737-ABBS72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 1706 BP; 432 A; 446 C; 458 G; 370 T; 0 other;

```

alignment_scores:

```

Quality: 917.00 Length: 478
Ratio: 2.866 Gaps: 10
Percent Similarity: 66.946 Percent Identity: 40.377

```

alignment_block:

```

US-09-701-572-2 x ABL07411 ..

```

Align seg 1/1 to: ABL07411 from: 1 to: 1706

```

42 ThrSerProSerArgThrIle...TyrSerAspArgPheIleProSerAr 57
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
354 ACCAGCCCTCTTAAGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 403
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
57 gSerAlaSerLysPheAlaLeu.....PheAspIleAsnThrProThg 72
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
404 GGGGGCTACCAACTTGTAGTACACACTTGTGTGTAACAAGAAGACTCG 453
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
72 LuGlyArgAspAspSerSerSerAlaTyrThr..... 83
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
454 GCGATAGTCGATGAGAGAAACAGACAGGACCACTCGACAAACAGCAAC 503
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
84 .....Le 84
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
504 GAGAGCAATGTCAGGCTCGCTCCAGAGGCGAGCGGAGCAAACTCAT 553
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
84 uLeuArgThrAlaLeuPheGlyProAspValAlaGly..... 96
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
554 CTCTGAAGTGGCCAGGTCGGTACTCCAAGGGCGGCGCATTTGTGCT 603
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
97 .....ProValThrProGluLysThrAspSerProSerMet 108
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
604 ACCAAACAGAGCTCCGCTGCTCCAGAAACACACACAACTCCCTGAG 653
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109 ThrLeuProAsnArgAsnIlePheArgTyrLysThrGluThrArgGlnSe 125
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
654 GTCTGTG.....TACTCATTAATACACCACTATATC 682
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
125 rMetHisSerLeuSerProPheMetAspAspPheValProGlyVala 142
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
683 CACAAAGAGTGGCTCA..... 698
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
142 snHisSerProValLysAlaProArgLysValProArgSerProTyrLys 158
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
699 .....CGCTATATATACCCACACACACACACAGGAG 725
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
159 ValLeuAspAlaProAlaLeuGlnAspAspPheTyrLeuAsnLeuValas 175
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
726 ATTCTGAGTGCACCTGATTTTATTAACGATTACTTAAATCTTATGGA 775
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
175 pTyrSerSerHisAsnValLeuAlaValGlyLeuGlyAsnGlyValTyrL 192
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
776 TTGGAGTGGCAGCATATATAGTGGCTGTGGCTGGAGTGGCTGAT 825
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
192 euTrpAsnAlaCysSerSerLysValThrLysLeuGlyAspLeuGlyVal 208
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
826 TGTGGAAGCAGACACCGAAATATCGACAGCTTACGGAGTTTGAGAGAG 875
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
209 AspAspCysValCysSerValGlyTrpAlaGlnArgGlyThrHisLeuAl 225
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
876 GGCGACTACGAGGCTCGCTATCGTGAATCCAGAGGAGGAGGAGACTTGC 925
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
225 aValGlyThrAsnAsnGlyLysValGlnIleTrpAspAlaAlaArgCysL 242
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
926 CATCGGCAACAGCACCGGCTGGGAGCTGTGGAGCTGCTCCAAAGTGA 975
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
242 yLysIleArgSerMetGluGlyHisArgLeuArgValGlyAlaLeuAla 258
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
976 AGCGTCTCGAGTGTGATGATGACACAGTCCGAGTGGATTCCTTGCC 1025
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
259 TrpSerSerSerLeuLeuSerSerGlyLysArgAspLysAsnIleTyrG 275
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1026 TGAAGTCAATTCCTGTTCTCTGGGAGCGCGGATGGACACATTGTC 1075
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
275 nArgAspIleArgThrGlnGluAspPheValSerLysLeuSerGlyHisL 292
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1076 CCAGAGTGTGCTGACGTCAGCAAGCTTCCACATTTCCGAGACACA 1125
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
292 ySerGluValCysGlyLeuLysTrpSerTyrAspAsnArgGluLeuAla 308
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

1126 GCGAGAGGTTTGGCGCCTAAAGTGTCCACGAGATTTCATATTGGCT 1175
309 SerGlyGlyAsnAspAsnLysLeuPheValTrpAsnGlnHisSer..... 323
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1176 AGCGGAGCAACAGCAAACTGTGTGATGTGTGTCGGGCGCACCGGTGG 1225
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
324 .....ThrGlnProValLeuLysTyrCysGluHisThrAla 336
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1226 CGTGGAGACTGCCACCGATTCCTTCGACAAATTCACAGCATTCAACCTG 1275
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
336 lavalysAlaIleAlaTrpSerProHisLeuHisGlyLeuLeuAlaSer 352
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1276 CAGTGTGCTGCTGTGGCTGTGCTCCGCAACCAAGTACTGTAGCTCT 1325
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
353 GlyGlyGlyThrAlaAspArgCysIleArgPheTrpAsnThrThrTras 369
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1326 GGAGGCGGACCGCGGATCGCTGCATCAAGTTCGAAATGGAACAATGG 1375
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
369 nSerHisLeuSerCysMetAspThrGlySerGlnValCysAsnLeuValT 386
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1376 CACTTATATGAATCCGCGGACTCCAAAGTCGACAGCTGTCTGTGCTCT 1425
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
386 rPserLysAsnValAsnGluLeuValSerThrHisGlyTyrSerGlnAsn 402
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1426 TTCTCGGCACTACAGAGAGCTGATCTGTGGAGTGTTCGTAACAC 1475
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
403 GlnIleIleValIleTrpArgTyrProThrMetSerLysLeuAlaThrLeu 419
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1476 CAACTGACCAATTTGGAATACCCCAACAATGTGTGAAGCAACCATTTGAC 1525
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
419 rGlnHisThrTyrArgValLeuTyrLeuAlaIleSerProAspGlyLys 436
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1526 TGGACACACGACGACGATCTCCAGATGCCATGTCTCCGAGCGGACGA 1575
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
436 hrIleValThrGlyAlaGlyAspGluThrLeuArgPheTrpAsnValPhe 452
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1576 CAGTATCAGCGCGGAGCTGATGAATACCTGCGTCTTGGACAGCTGCT 1625
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
453 ...ProSerPro.....LysSerGlnAsnThrCysLeuGlu 464
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1626 GCTCCGATCGTGTGGCGCTCAAGAGGAGGAGTTCGACGACGAG..... 1670
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
464 eGlyAlaLeuSerLeuGlyArgThrThrIleArg 475
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1671 .GGCAACAGAGGCGGTTCGACAGAGCAATCCGT 1703
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
seq_name: /stids1/gcgdata/hold-geneseq/geneseqn-emb1/NA195.DAT:AAQ96099
seq_documentation_block:
ID AAQ96099 standard; cDNA; 1767 BP.
XX
AC AAQ96099;
XX
DT 23-JAN-1996 (first entry)
XX
DE Sequence encoding rat cell cycle protein p55CDC.
XX
KW Cell cycle; protein; p55CDC; antibody; chemotherapy; modulation;
XX inhibition; growth; ss.
XX
OS Rattus rattus.
XX
XX
FH Key location/Qualifiers
FT CDS 193..1692
FT /tag= a
FT /product= Cell cycle protein p55CDC.
XX
PD 17-AUG-1995.
XX
PF 13-FEB-1995; 95WO-US01806.
XX

```

PR 14-FEB-1994; 94US-0195730.
XX
PA (AMGE-) AMGEN INC.
PA (REGC) UNIT CALIFORNIA.
XX
PI
XX Weinstein J;
WP1: 1995-293119/38.
DR P-P-SDB; AAR78656.
XX
PT Mammalian cell cycle protein p55CDC and DNA sequences encoding it -
PT and methods of modulating cell division using cpds, effecting p55CDC
PT or associated complexes useful in e.g. chemotherapy";
XX
PS Claim 1; Page 51-54; 87pp; English.

The p55CDC polypeptide is involved in mammalian cell division. DNA,
proteins and antibodies derived from the polypeptide can be used to
modulate cell cycle activity e.g. in chemotherapy to inhibit the
growth of tumour cells.

Sequence 1767 BP; 413 A; 487 C; 487 G; 380 T; 0 other;

```
alignment_scores:
  Quality: 900.00      Length: 498
  Ratio: 2.711        Gaps: 9
Percent Similarity: 66.667  Percent Identity: 39.157
```

```
alignment_block:
US-09-701-572-2 x AAQ96099 . .
```

Align seg 1/1 to: AAQ96099 from: 1 to: 1767

```

8 AsnProProProIothSerThyValAlaArgspAsnSerProProProIuPr 24
256 AATGACACGATTCCTGCTGTGCAGGACGAAGCAAGCAAGCCACAGGCC
24 oSerProGluSerLeuArgHisValSerArg..... 34
106 AGCCCCCTGGCCATGAGGGGGCCCAACAGATCACACAGGCCGGTGGA 3555
35 .....MetIleasnSerAsnHisTyrThrSerPro 44
356 CCCCAGGCGCAATCCTGGCAATCTAATCTTAAGTTCAACACCCACCTT 405
45 SerArgThrIleTyrSerAspArgPheIleProSerAlaGlySerAlaSerly 61
406 AGCAAACT...GGAAGTAGGGCTATATCCCCCAAGCTAGTGGTCCCA 4232
61 s.....PheAlaLeuPheAspIleAsnThrProThrGluG 73
453 AATGAGAGTGGCCAGGTTCTCTTGAGCAAGGAAACAGCGGAGAGCG 502
73 LysArgAspAspSerSerSerAlaTyrThrThrLeuLeuArgThrAlaLeu 89
503 GGGGTACGCCACCAAGAGAGAGCATCAGAAAGCCTGGGCTCGGAACTG 552
90 PheIlyProAspValAlaGlyProValThrProGluLysThrAspSerPr 106
553 AACGGCTTTTATGTG..... 567
106 oSerMetThrLeuProAsnArgAsnIlePheArgTyrLysThrGluThra 122
568 .....GAGAAAGCCAAAGTCTCTCAGGCTCAGTGGAAACCTC 604
123 rglInser.....MetHisSerLeuSerProPheMetAspAsp 133
605 AGAATGCCCCAGAAGGCTACACAGAACAGATGTAAGATCTCTACAGCCAG 65
136 AspheValProGlyValAsnHisSerProValLysAlaProAlaGlyVal 152
655 AAAGCCCAAGGCTGGC.....TTCAGTGGAAAGGCTGACGATACAT 699

```

152	1	proabserprotyllysvalleuaspalaproala	leuglaspasp	169
153				
696	TCCTTCCTCCGACAGACAGATTTCTTATGCCCTCGAATCCGGAACTACT			745
169	hetyrleuasnleuvalasprpserseerhisasnvalleuvalaValgly			185
746	actacgaactgattctgattggagctctggaatgattatgactgtggca			795
186	leuglyasnCysValtylLeutPrasnAlacyserserleuValLthrly			202
796	ctggcgaacacagctgtactgtatggaaacgtgtgtccggtacatctggca			845
202	slenucysleu...glylaaspaspCysValCysSerValglyTrpA			218
846	gctgttgcaaatggacagacgctggggactacatattcatccgtgctggca			895
218	laqlnarglyThrhislleuAlaValglyThrasnanglylysValgln			234
896	tcgaagagggcgaacactcgtctgtggccacacactaaagctgagctgacg			945
235	lhetppapAlaAlaargCyslyslysIleargsermetglnghislar			251
946	ctatggagatgtccagacacagaaacgcttggaaacatgaccacacactc			995
251	gleuargValglyAlaAlaLeuAlatrpSerSerleuLeuSerSerolyg			268
996	tcctcgagaaacgtcccgagttggaaacagctaatatcctgcaagtgtt			1045
268	lyargAspIysasnIlelyrglnargAspIleargThnglnIasphe			284
1046	cacgaactggccacacatccacacacagatgttcgagtacgaaacacact			1095
285	valserIylsleuserGlyhislysergluValCysglyleuIysTrpse			301
1096	gtggccacactgagggccatagccacagaaagtatgtggctggcctggcc			1145
301	rtyrAspAsnargIuIleuAlaSerGlyIysAsnAspAsnlyslleuphey			318
1146	ccacagatggacacatctggcgaagggtggcgaatgataactgtgcaacg			1195
318	alTrp.....asnghisIserThglnProvalleuIysTrp			330
1196	tetggcctagtgtcctctgagaaatggctggcgtccctggacacatctc			1245
331	CysGlnhisThralaAlaVallysAlaIleatrpSerProhisIlehi			347
1246	actcaacactcaagagctctcgaagcctgtgtcattgctccctggcactc			1295
347	sgIyIeuAlaIaserglyGlyIThralaAspArgCysIleargPheT			364
1296	caatattccctggcaacagagagagatracacagacacatntcccatTT			1345
364	rPasnThrThrasnSerhisIleuserCysmetAspThrGlySerGln			380
1346	ggaaagctgtgctgtgagcctctgagagctgtgagatgtgcattccag			1395
381	ValCysAsnleuValTrpserIysAsnValasnghIysuIysSerThhi			397
1396	gtgtgtctcatctctgtgtctccacataaaggagctcatctcaggcca			1445
397	sgIyTrserGlnAsnGlnIleIeValTrpArgTrpProhImetSerL			414
1446	tgacttggcccaacacacagctgtatttgggaagAACCAACCAATGGGCA			1495
414	ysLeuAlaThrleuThrghlyhisThrTyArgValleuTyIeuAlaIle			430
1496	agctggcacaagctcaaaagctcacaacgcccgtctgctgctcacaatg			1545
431	serProAsnGlyGlnThrIleValThrhgIValagIyAsnGlyThrLeuAr			447
1546	agctccacagagggccacacatgtgactctgacagacgccgatgacactctgcg			1595


```

61 sphenalaleupheaspilleasnthrProthrGluGlyArgaspaspsers 78
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
377 GATGGAGGTGGCAGCTTCCTCTGAGCAGAGAACCACTGTGAAAAACA 426
78 eSerAlaTyrThrThrLeuLeuArgThrAlaLeuPheGlyProAspVal 94
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
427 GCCAG..... 431
95 AlaGlyProValThrProGlyLysThrAspSerProSerMetThrLeuR 111
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
432 .....ACGCCCAACAGAGAGACATCAGAAAGCCTGGGCTTT 469
111 oasnargasn.....IlePheArgTyrLysT 120
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
470 GAACCTGAACGGTTTGTATGTAAGAGAACCAAGATCCTTGCGCTCAGT 519
120 hrGluThrArgGlnSer.....MetHisSerLeuSerProPhe 132
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
520 GAAACCAACAAATAGCGCAGAGGTTATCAGAACAGACTGAAAGTACTC 569
133 MetAspAspAspPheValProGlyValAsnHisSerProValLysAlaPr 149
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
570 TACAGCCAAAGGCCACTCTGGC.....TCCAGCCGAGAGACCTG 610
149 oArgTyrValProArgSerProTyrLysValLeuAspAlaProAlaLeuG 166
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
611 CCGTTACATTCCTTCCTGCGCAGACCGTATCTCTGATCGCCTGAAATCC 660
166 InAspAspPheTyrLeuAsnLeuValAspTrpSerSerHisAsnValLeu 182
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
661 GAAATGACTATTAACCTGAACCTTGATGAGTTCGGAGATCGGAATGTA 710
183 AlaValGlyLeuGlyAsnGlyValLysThrLeuThrPasnAlaGlySerSer 199
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
711 GCGTGGACCTGGACACAGCTGTGACTGTGAGTGCACACTCTGGTGA 760
199 sValThrLysLeuGlyAspLeu...GlyValAspAspCysValCysSerV 215
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
761 CATCTGAGAGCTTTGCAATGAGCAGCCTGGGCAATATATATACCTCTG 810
215 aGlyTyrPalaGlnArgGlyThrHisLeuAlaValGlyThrAsnAsnGly 231
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
811 TGGCCTGATCAAAAGAGGCAACTCTGCTGGTGGCAGCAGCAGCTCT 860
232 LysValGlnIleTrpAspAlaAlaArgCysLysLysIleArgSerMetG 248
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
861 GAGGTGCAGCTATGGATGTGCAGCAGCAAGAACGGCTTCGAAATATAT 910
248 uGlyHisArgLeuArgValGlyAlaLeuAlaTrpSerSerSerLeuLeu 265
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
911 CAGTCACCTGCGCCGAGTGGCTCCCTAAGCTGAACAGCTATATCTCT 960
265 eSerGlyGlyArgAspLysAsnIleTyrGlnArgAspIleArgThrGln 281
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
961 CCAGTGGTTCAGCTTCGCAACATCCACACCATGATGTCGGGTACCA 1010
282 GluAspPheValSerLysLeuSerGlyHisLysSerGlyValLysGlyLe 298
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1011 GAACACCATGTGGCCACACATGATGGCCACACCGAAGATGTGGGCT 1060
298 uLysTrpSerTyrAspAsnArgGlyLeuAlaSerGlyLysAsnAspAl 315
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1061 GCGCTGGGCCCCAGATGAGCAGCATTTGGCAGTGGTATATGATACT 1110
315 ySLeuPheValTrp.....AsnGlnHisSerThrGlnProVal 327
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1111 TGGTCAATGTGTGCTAGTCTCTCGAGAGGGGTGGCTGGCTCTCTG 1160
328 LeuLysTyrCysGlnHisThrAlaAlaValLysAlaIleValTrpSerPr 344
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1161 CAGACATTCACCCACATCAAGAGGCTGTCAAGAGCCGTAGATGTGTGCC 1210
344 OHISLeuHisGlyLeuAlaSerGlyGlyGlyThrAlaAspArgCysI 361

```

```

1211 CTGGCAGCTCCATATGCTCTGGCACACAGAGGGGCGACCATGATCGACACA 1260
361 leArgPheTrpAsnThrThrThrAsnSerHisLeuSerCysMetAspThr 377
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1261 TTCGATCTGTGAATGTGTGCTGTGGGCGCTGTGTGAGTCCCGGTAGTGC 1310
378 GlySerGlnValCysAsnLeuValTrpSerLysAsnValAlaGlnLeuVal 394
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1311 CATTCGAGGTGTGTCTCATCTGTGTCTGCCCATTCACAGAGGCTCAT 1360
394 lSerThrHisGlyTyrSerGlnAsnGlnIleIleValTrpArgTyrProT 411
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1361 CTCAGGCGCATGGCTTGACAGAACCAAGCTATTTGAGATACCA 1410
411 hrMetSerLysLeuAlaThrThrGlyHisThrTyrArgValLeuTyr 427
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1411 CCATGGCCAGGTGGCTGAACCTCAAGGTCACACATCCGCGGCTCGAGT 1460
428 LeuAlaIleSerProAspGlyGlnThrIleValThrGlyAlaGlyAspG 1444
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1461 CTGACCATGAGCCAGATGGGCGCCACAGTGGCATCCGACAGCAGATGA 1510
444 uThrLeuArgPheTrpAsnValPheProSerProLysSerGlnAsnThrG 461
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1511 GACCTGAGGCTATGGGCGCTTTTGTAGTGTGACCTCGCGGCGGCGG 1560
461 luserGlu 463
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1561 AGCGGGAG 1568

```

seq_name: /sids1/gcdata/hold-geneseq/geneseg-emb1/NA2001B.DAT.ABL07410

seq_documentation_block:

```

ID ABL07410 standard; CDNA; 3852 BP.
XX
AC ABL07410;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 16712.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmacetical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
DR P-PSDB; ABB63307.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 16712; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of

```

CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL01840-ABL01849), expressed DNA
CC sequences (ABL01840-ABL01849), and the encoded proteins
CC (AB057373-AB057382).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp://wipo.int/pub/published_sequences.

XX Sequence 3852 BP; 1092 A; 849 C; 881 G; 1030 T; 0 other;

alignment_scores:

Quality: 859.00 Length: 519
Ratio: 2.659 Gaps: 13
Percent similarity: 62.235 Percent identity: 37.187

alignment_block:
US-09-701-572-2 x ABL07410 ..

Align seg 1/1 to: ABL07410 from: 1 to: 3852

42 ThrSerProSerArgThrIle...TyrSerAspArgPheIleProSerAr 57
1354 ACCAGCCCTCTACAGACGCGAGAGGCGAGATGCGCTTATTCGCAATCG 1403
57 GSerIaSerLysPheAlaLeuPheAspIleAsnThrProThrGluGly 74
1404 GCGCGCTACCACTTGTGAGTTAGCACACTT... 1434
74 rAspAspSerSerSerSerAlaTyrThrIleLeuAlaArgThrAlaLeuPhe 90
1435CTGGTAAGGATTCCTCACTCAATC 1455
91 GlyProAspValAlaGlyProVal...ThrPro.. 100
1456 CCTCCA.....TCCCGCCCTGTAAGGCGATCCTTATTCAGACCCCTTA 1499
100 100
1500 AGAATTATTGACACACGTCGACATCTATTAGTGAACAAAGACTCCGGCG 1549
101GlyLysThrAspSerProSerMetThrIleu 110
1550 ATTAAGTCGATGAGAGAGACGACAGCCCTCGACGACAGCAGCAGAG 1599
111 ProAsnArgAsnIlePheArgTyrLysThrGluThrArgGlnSerMetH1 127
1600 AGCAATGTCAGGCTTCGCTCACAAGGCGCAGCGCAAACTCATCTC 1649
127 sSerLeuSerProPheMetLaspAsp..... 135
1650 TGAAGTGGCCAGTGGTGGTCAAGGCGGCGCATTTTGTGCTACG 1699
136AspPheValProGlyValAlaAsnHisSerProValLys... 147
1700 AAAACAAAGGCTCCGCTGCTCCAGAACACACACAAATCCCTGGAAGTTC 1749
148AlaProArgLysVal 152
1750 GTGTACTCATTAAGACACCCATATCCCAAGAGTGGCTCAGCTAATAT 1799
152 lProArgSerProTyrLysValLeuAspAlaProAlaLeuGlnAspAsp. 168
1800 ACCGACACATCCGAGAGGATTCGATGACACCTGATTTATTATTAAGATT 1849
168 168
1850 ACTGATAGTGCAGAAACTTTCTTTCTTTTCTTCTTCTTCTTCTTCTTCT 1899
169 PheTyr..LeuAsnLeuValAspTrpSerSerHisAsnValLeuAlaVal 184
1900 TTCAGATTTAATCTTATGATGTGAGTGGAGTCCGACATATTAAGTGGCTG 1949

185 GlyLeuGlyAsnGlyValTyrLeuThrPheAsnAlaCysSerSerValThr 201
1950 GCGTGGGAGGAGTGGTCTTATTTGTGGAGCGACAGCCGGAATATCGA 1999
201 rLysLeuGlyAsnGlyValAlaAspAspCysValLysSerValGlyTrpA 218
2000 GCACCTTACGAGATTTAGGAGGAGGCGGACTACGACAGGCTGCTTCTGGA 2049
218 lArgIaArgGlyThrHisLeuAlaValAlaGlyThrAsnAsnGlyLysValGln 224
2050 TCCAGAGAGGCGACAGATATTGCTCCATCCGCAACAGACAGCGTGGCTGAG 2099
235 lLeuTrpAspAlaAlaArgCysLysLysIleArgSerMetLysGlyHisAr 251
2100 CTGGGAGTGGCTCCAAAGTAAAGCTGCTCGAGTGAATGATGATGACAG 2149
251 gLeuArgValGlyAlaLeuAlaTrpSerSerSerLeuLeuSerSerGlyG 268
2150 TGCCGAGTGGGATCTTGGCTGGAGTCACTTCCTGCTTCTGCTGCA 2199
268 lYArgAspLysAsnIleTyrGlnArgAspIleArgThrGlnGluAspPhe 284
2200 GCGGAGATGGACCATTTGTCACACAGATGCGGTGACGTCGACGACAG 2249
285 ValSerLysLeuSerGlyHisLysSerGlyValCysGlyLeuLysTrpSe 301
2250 CTTCACATTTGCTCCGACACAGCAGAGAGTGGCTGCTAAAGTGTGTC 2299
301 rTyrAspAsnArgGlyLeuAlaSerGlyGlyAsnAspAsnLysLeuPheV 318
2300 CAGGATTTCAAGATTTGGCTACGCGAGGACAGCAACATCTGTGTAATG 2349
318 alTrpAsnGlnHisSer.....ThrGlnProValLeu 328
2350 TTGTGTCGCGCGGCGGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2399
329 LysTyrCysGlnHisThrAlaAlaValLysAlaIleAlaTrpSerProH1 345
2400 AAATTCAGCGACATCAAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGT 2449
345 sLeuHisGlyLeuLeuAlaSerGlyGlyGlyThrAlaAspArgCysIleA 362
2450 GCACSCAAGTCTCTACCTCTGAGGCGGCGGCGGCGGCGGCGGCGGCGG 2499
362 rGpHeTrpAsnThrThrThrAsnSerHisLeuSerCysMetAspThrGly 378
2500 AGTCTGGAATGTGAACAATGACACTTAAATGAATCCGTGACTCCAG 2549
379 SerGlnValCysAsnLeuValTrpSerLysAsnValAsnGlnLeuValSe 395
2550 TCGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2599
395 rThrHisGlyTyrSerGlnAsnGlnIleIleValTrpArgTyrProThrM 412
2600 TCGCATAGGTTTGTCTACACACACACTGACCATTTTGGAAATACCCACAA 2649
412 eSerLysLeuAlaThrLeuThrGlyHisThrTyrArgValLeuTyrLeu 428
2650 TGTGAAGCAAGCCGATTTGACTGTGACACACAGTCACAGATTCACAGATG 2699
429 AlaIleSerProAspGlyGlnThrIleValThrGlyAlaGlyAspGlyuth 445
2700 GCCATGCTCTCGGAGGCGACAGATGATGATGATGATGATGATGATGATG 2749
445 rLeuArgPheTrpAsnValPhe...ProSerPro.....LysS 457
2750 CTGCGTCTTTGGAATGCTTCCGCTCCGATCCGCTTGGCTCCAGAGAG 2799
457 eArgIaAsnThrGlySerGlyIleGlyAlaLeuSerLeuGlyArgThrThr 473
2800 CAGTTTGCAGCAGCAG.....GGCAACAGAGCGTGTTCGACAGAGC 2843
474 lLeArg 475

|||||
2844 ATCCGT 2849

seq_name: /SIS1/gcgdata/hold/geneseq/geneseqn-emb1/NA2000.DAT:AAc46256

seq_documentation_block:

ID AAC46256 standard; DNA; 2697 BP.

XX AAC46256;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 49475.

DE Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

XX EPI033405-A2.

XX PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134236.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134321.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139753.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 02-JUL-1999; 99US-0142154.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 20-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 21-JUL-1999; 99US-0144632.

PR 21-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.

PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148688.

PR 16-AUG-1999; 99US-0149368.

Wed May 29 14:59:19 2002

us-09-701-572-2.rng

Page 21

652 GAAGCAGTTTCGACCAGCAG.....GCCAACAGAGCGTGTTCGAC 695
472 hThtlIeAtg 475
:::|||||
696 AGAGCATCGT 706

OM of: US-09-701-572-2 to: EST.* out_format: pfs
Date: May 29, 2002 10:12 AM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frmt-p2n.model -DEV=xlh
-O=/cgn2_1/US09701572/runtat_29052002_094114_29840/bpp_query.fasta_1.538
-DB=EST -QMT=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.500 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -GAPOP=6.000
-GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09701572_ECGN1.1.6618
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPHY-WAIT -THREADS=1

Search information block:

Query: US-09-701-572-2
Query Length: 475
Database: EST.*
Database sequences: 13736207
Database length: 1841457050
Search time (sec): 1579.010000

```
score_list:
Sequence      Strd Orig      ZScore      EScore      Len      Documentation
gb.est1:BG584091 + 1148.00 1787.35 2.6e-90 759 1*BG584091 EST485851 MHAM Mediacg
gb.est2:BI270288 + 1122.00 1748.69 3.7e-88 630 1*BI270288 NF006080BFLP1065 Devel
gb.est2:BG044933 + 1111.00 1731.10 3.5e-87 648 1*BG044933 sa35d02.y1 GM-cl059 G
gb.est2:BG444274 + 1102.00 1712.15 4.0e-86 987 1*BG444274 GA_EB0023N11f Gossyp1
gb.est1:AM050073 + 1001.00 1557.74 1.7e-77 710 1*AM050073 EST273990 tomato calli
gb.est1:BI456001 + 976.00 1515.84 3.5e-75 833 1*BI456001 603170445F1 NCI_CGAP_M
gb.est2:BG760896 + 917.00 1423.78 4.7e-70 833 1*BG760896 602717081F1 NIH_MGC_45
gb.est2:BI655256 + 902.50 1401.69 7.9e-69 786 1*BI655256 603299431F1 NIH_CGAP_I
gb.est2:BI904176 + 895.50 1389.56 3.3e-68 913 1*BI904176 60316804F1 NCI_CGAP_I
gb.est2:BM462011 + 895.00 1388.89 3.1e-68 860 1*BM462011 ACENCOUPR_6418306 NIH
gb.est1:BI895812 + 892.00 1390.39 3.4e-68 501 1*BI895812 EST262255 tomato calli
gb.est2:BI895812 + 876.50 1359.82 1.7e-66 863 1*BI895812 603236792F1 NIH_CGAP_M
gb.est2:BI895812 + 868.00 1346.15 9.8e-66 888 1*BI895812 60158251F1 NIH_MGC_7
gb.est2:BE795056 + 865.00 1343.21 1.4e-65 873 1*BI895812 602917131F1 NCI_CGAP_I
gb.est2:BI155551 + 862.00 1338.59 2.6e-65 756 1*BI904107 60316667F1 NCI_CGAP_I
gb.est2:BI904107 + 853.00 1314.63 5.6e-64 1781 1*AK011723 Mus musculus 10 days
gb.hic:AK011123 + 851.00 1316.33 4.5e-64 1168 1*AK011123 Mus musculus 10 days
gb.hic:AK011131 + 839.00 1303.08 2.5e-63 718 1*AM476563 uq77g09.y1 NCI_CGAP_Lu
gb.est1:AM476563 + 834.50 1293.58 8.3e-63 888 1*BI554730 603236692F1 NIH_CGAP_M
gb.est2:BI554730 + 825.00 1284.53 2.7e-62 533 1*AM065517 614059E06.y1 614 - roc
gb.est1:AM065517 + 811.00 1256.06 1.0e-60 939 1*BE263129 601144853F2 NIH_MGC_18
gb.est2:BE263129 + 801.50 1239.81 8.2e-60 1056 1*AL569765 AL569765 LIT_NFL006_1
gb.est1:AL569765 + 783.00 1218.12 1.3e-58 557 1*BE329969 sc01910.y1 GM-cl040 G1
gb.est2:BE329969 + 781.00 1215.04 2.0e-58 554 1*AL770538 6005054A05.x2 606 - Est
gb.est1:AL770538 + 751.00 1162.27 9.1e-55 974 1*AL526768 AL526768 LTI_NFL003_NH
gb.est2:AL526768 + 751.00 1162.27 9.1e-55 974 1*BE003530 EST432028 KVI Mediacg
gb.est1:BE003530 + 742.00 1148.30 1.0e-54 898 1*BE561172 601344895F1 NIH_MGC_8
gb.est2:BE561172 + 740.50 1149.90 8.4e-55 636 1*BI067521 BI067521 NIBB Mochli_1
gb.est2:BI067521 + 738.00 1148.37 1.1e-54 533 1*BI736659 SB1103.3.y1 GM-cl012 G1
gb.est1:BI736659 + 726.50 1129.17 1.2e-53 571 1*BM004797 d433610.y1 NICHD XGC
gb.est2:BM004797 + 725.50 1127.76 1.4e-53 563 1*BI420244 LjWHE5754912r Lotus Jaf
gb.est2:BI420244 + 718.50 1115.19 5.6e-53 548 1*BE032617 132042 MRC LPIC Sus s
gb.est1:BE032617 + 717.50 1113.19 7.2e-53 564 1*AL484860 EST242121 tomato ovary
gb.est1:AL484860 + 708.00 1096.21 8.2e-52 804 1*BI467188 BOG0800TR BOG6 Brassic
gb.gss:BI467188 + 706.00 1095.20 9.3e-52 878 1*BI456456 HVSBE0019N04f Hordeu
gb.est1:BI456456 + 702.50 1084.68 1.0e-51 699 1*AL506441 Hordeu vullg
gb.gss:AL506441 + 701.00 1088.32 2.3e-51 1051 1*AL336771 77 end of clone AR0A
gb.gss:PM021801U + 700.00 1080.92 5.8e-51 1020 1*AL433472 77 end of clone BB0A
gb.gss:CNS0782Y
```

```
gb.est1:AV112537 + 698.50 1078.69 7.7e-51 1009 1*AV112537 AV112537 Mus muscu
gb.est1:AL138327 + 690.00 1071.30 2.0e-50 601 1*AL138327 DRT2p5470057.r1 547
gb.est2:BE286879 + 688.00 1063.58 5.4e-50 896 1*BE286879 601092614F1 NCI_CGA
gb.est1:BE204118 - 684.00 1065.92 4.0e-50 556 1*BE204118 EST396794 KVO Medic
gb.gss:BIH573977 - 684.00 1060.37 8.1e-50 686 1*BIH573977 BOGER48TF BOGE Bras
```

seq_name: gb.est2:BG584091

seq_documentation_block:

LOCUS BG584091 759 bp mRNA linear EST 11-APR-2001
DEFINITION BG584091 MHAM Mediacg truncatula/Glommus versiforme mixed EST
library CDNA clone pMHAM-14B17 5' end, mRNA sequence.
ACCESSION BG584091 GI:13599155
VERSION BG584091.1
KEYWORDS EST.
SOURCE Mediacg truncatula/Glommus versiforme mixed EST library.
ORGANISM Mediacg truncatula/Glommus versiforme mixed EST library.
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 759)
Harrison,M.J., Liu,J., Town,C.D., Van Aken,S., Utterback,T., Cho,J.
and Fraser,C.M.
ESTs from roots of Mediacg truncatula after colonization with
Glommus versiforme, 2001
TITLE Unpublished (2001)
JOURNAL Contact: Harrison M.J.
COMMENT Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73401
Tel: 580-223-5810
Fax: 580-221-7380
Email: mjharrison@noble.org
Noble EST name: N3795616 TIGR sequence name: MTDBM09TK More
information is available at: <http://www.medicago.org>
Seq primer: Skm0d (CTA gaa gtg gat cc).
Location/Qualifiers

FEATURES

source

```
1..759
/organism="Medicago truncatula/Glommus versiforme mixed EST
library"
/cultivar="Medicago truncatula genotype A17"
/db_xref="taxon:119092"
/clone="pMHAM-14B17"
/tissue="MHAM"
/clone_lib="MHAM"
/clone_type="roots colonized with Glommus versiforme"
/deriv_stage="Roots harvested at 10, 17, 22, 31 and 38 days
post-inoculation with Glommus versiforme. The library was
made from a mixture of RNA from each of these stages."
/lab_host="E. coli strain XL0R"
/notes="Vector: pBluescript SK-. Site 1: EcoRI; Site 2:
XhoI; CDNA was prepared from polyA+ enriched RNA from
roots harvested at 10, 17, 22, 31 and 38 days
post-inoculation with Glommus versiforme. The CDNA was
directionally ligated into the Unizap Xk vector from
stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing CDNA inserts were excised
from the recombinant lambda-zap phage using Ex-assist
helper phage and propagated in XL0R cells."
```

BASE COUNT 187 a 165 c 185 g 222 t

ORIGIN

alignment_scores: Length: 230
Quality: 1148.00 Gaps: 3
Ratio: 5.148
Percent Similarity: 96.957 Percent Identity: 96.087

alignment_block: US-09-701-572-2 x BG584091 ..

Align seg 1/1 to: BG584091 from: 1 to: 759

```
1 MetAspGlyTThGlyAsnArgAsnProProThrSerThValArgAs 17
|||||
73 ATGGACGAAACCGGATATCGAATACACACGACCTTCACCGTGAGAG 122
```

```

17 pasSerProProgluProSerProgluSerLeuArgHisValSerA 34
123 TAAATTCACACCGCCGAGCATACCGGAGAGCTCCGCCCATGTAAAGCC 172
34 rgeMellLeasSerAsnHisTyrThrSerProSerArgThrIleTyrSer 50
173 GTATGATCAACAGCACCATTTACACCTCACCTTCTCGAACATCTACTCC 222
51 AsparGhellProSerArgSerAlaSerLysPheAlaLeuPheAspIle 67
223 GATAGGTGATCCGAGTACATCTGCTTCAAAATGCGCTTGTGATAT 272
67 easnThrProThrGluGlyArgAspAspSerSerSerAlaTyrThrL 84
273 CAATACCTCCAGCAGAGCAGCATGATATTCACAGCGCTTATACGACTC 322
84 euleuAtgThrAlaLeuPheGlyProAspValAlaGlyProValThrPro 100
323 TTCTTACAAAGCGGCTGTTGAGACCGGATGTCAGGTCGCGTTACGCG 372
101 GluThrAspSerProSerMetThrLeuProAsnArgAsnIlePheArg 117
373 GAGAAACGACCTCGCGGTGATGACATTTCCGAAATAGGAATATTTTAg 422
117 gTyThrGluThrArgLysSerMetHisSerLysLeuSerProPheMetA 134
423 GTATTAAGACGAGACGAGACGATCCATGCACTGCTTGGCGCTTATAG 472
134 spAspAspPheValProGlyValAsnHisSerProValLysAlaProArg 150
473 ATATGATTTTGTCTCGTATTAATCATAGTCGGGTAAAGGCTCTTAG 522
151 LysValProArgSerProTyrLysValLysAspAlaProAlaLeuGlnAs 167
523 AAGTTCCCTCGATCGCCTTATAGCTTTTGATGACACCTGCTTGCACAA 572
167 pasPheTyrLeuAsnLeuValAspTyrSerSerHisAsnValLysAla 184
573 TGATTTTATCTGAATCTGTAGATGTTGCTTACACAAATGTTGGCTG 622
184 alGlyLeuGlyAsnGlyValTyrLeuTyrAsnAlaCysSerSerLysVal 200
623 TTGGTTGGGTAACTGTGTTTATCTCTGGAATGCTTGAAGCAGCAAGGA 672
201 ThrLysLeuCysAspLeuGlyValAspAspCysValCysSerValG 216
673 ACTAATATATGATTTGGGGGGGTGATGATGAGGGGTTGTTCTGTGTTG 722
216 yTTPAlaGlnArgGlyThrHisLeuAlaValGly 227
723 TTGGGCTCAACGCGGTACTCATCTGCTGTGGA 756
seq_name: gb_est2:BI270288
seq documentation block:
LOCUS BI270288 630 bp mRNA linear EST 18-JUL-2001
DEFINITION NE006A08FL1065 Developing flower Medicago truncatula cDNA clone
ACCESSION BI270288
VERSION BI270288.1 GI:14877742
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
REFERENCE 1 (bases 1 to 630)
AUTHORS Torres-Jerez, I., Scott, A. D., Harris, A. R., Gonzales, R. A., Bell, C. J.,
Flores, H. R., Imman, J. T., Weller, J. W., and May, G. D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula flower library

```

```

JOURNAL Unpublished (2001)
COMMENT Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380
Email: gdmay@noble.org
Insert Length: 630 Std Error: 0.00
Plate: 006 row: A column: 08
Seq primer: TCACACGAGAAACAGCATATGAC.
FEATURES
Location/Qualifiers
1..630
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NE006A08FL"
/clone_1ib="Developing flower"
/tissue_type="Developing flowers"
/dev_stage="Developmentally pooled. Contains a mixture of
very young, developing, fully-opened flowers and flowers
in early transition into pods."
/note="Vector: Lambda Zap; cDNA was prepared from polyA+
enriched, pooled samples of equivalent amounts of total
RNA from very young, developing, fully-opened flowers and
flowers transitioning into pods. The cDNA was
directionally ligated into the Uni-Zap XR vector
(Stratagene) and packaged using the GigaPack III Gold
packaging extracts. Phagmids containing cDNA inserts were
in vivo excised from the recombinant Uni-Zap XR vector
using Exassist helper phage and the E. coli strain
XLI-Blue MRF (Stratagene). Excised plasmids were plated
using SORR cells."
BASE COUNT 159 a 97 c 167 g 204 t 3 others
ORIGIN
alignment_scores:
Quality: 1122.00 Length: 209
Ratio: 5.420 Gaps: 0
Percent Similarity: 99.043 Percent Identity: 98.565
alignment_block:
US-09-701-572-2 x BI270288
Align seg 1/1 to: BI270288 from: 1 to: 630
112 AsnArgAsnIlePheArgTyrLysThrGluThrArgLysSerMetHis 128
|||||
3 AATAGGAATATTTTATGATATAGACGAGACGACATGCTCATGCTC 52
128 rLeuSerProPheMetAspAspPheValProGlyValAsnHisSerP 145
|||||
53 GCTTTCGCGCTTATGATATGATATTTTCTCCGATTAATCATATGTC 102
145 roValLysAlaProArgLysValProArgSerProTyrLysValLysAsp 161
|||||
103 CGGTAAAGGCTCTAGAGAGAGTCTCGATCGCCTTATAGGTTTGAT 152
162 AlaProAlaLeuGlnAspAspPheTyrLeuAsnLeuValAspTyrP 178
|||||
153 GCACCTGCTTGGCAAGATGATTTTATCTGAACTGTGATATGCTTTC 202
178 rHisAsnValLeuAlaValGlyLeuGlyAsnGlyValTyrLeuTyrAsn 195
|||||
203 ACACAAATGTGTGCTTGTGTTGGTAACTGTGTTATCTCTGGAATG 252
195 lacYSerSerLysValThrLysLeuCysAspLeuGlyValAspAspC 211
|||||
253 CTGTGACACAGCAAGGTAAATTAATGATTTGGGGGTGATGATTTG 302
212 ValCysSerValGlyTTPAlaGlnArgGlyThrHisLeuAlaValGly 228
|||||
303 GTTTGTTCTGTGTTGGGCTCAACGCGGTACTCATCTGCTGTGGAAC 352

```

228 TAsnAsnGlyValGlnIleTPAspAlaAlaArgCysLysIleA 245
 333 TAAcATGTAAGTTCAGATTGGATGCAGCAAGATGCAAGAAATAA 402
 245 rGSeMetGluGlyHisArgLeuArgValGlyAlaLeuAlaTrpSerSer 261
 403 GATCATGAGAGGGCCATCGGTACGTGCGGTGGCTGGCTGAGATTCA 452
 262 SerLeuLeuSerSerGlyGlyArgAspLysAsnIleTyrGlnArgAspI 278
 453 TCTCTTTGCTCTTCGTGAGCGGATAGATATTATATACACAGATAT 502
 278 eArGThrGlnGluAspPheValSerLysLeuSerGlyHisLysSerGlu 295
 503 ACGCACACAAAGATTTTGTAGTAACTGTCAGACACAAATCAANAG 552
 295 aLysGlyLeuLysTrpSerTyrAspAsnArgGluLeuAlaSerGlyGly 311
 553 TTTGTGGACTAGAGTGTATATATATACCGAGTTGGACTGTGAGAGA 602
 312 AsnAspAsnLysLeuPheValTrpAsn 320
 603 AATGACACAAATGTTGTTGGAT 629

seq_name: gb_est2:BG044933

seq_documentation_block:

LOCUS BG044933 648 bp mRNA linear EST 28-NOV-2001
 DEFINITION saas3d02.yl1 Gm-cl059 Glycine max cDNA clone GENOME SYSTEMS CLONE
 ID: Gm-cl059-939 5' similar to TR-09XED5 Q9XED5 CELL CYCLE SWITCH
 PROTEIN ; mRNA sequence.

ACCESSION BG044933

VERSION BG044933.1 GI:12492159

KEYWORDS EST.

SOURCE soybean.

ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.

REFERENCE 1 (bases 1 to 648)

AUTHORS Shoemaker R., Klein P., Vodkin L., Eipelting J., Coryell V., Khanna
 A., Bolla B., Matra M., Hillier L., Kucaba T., Martin J., Beck C.,
 Wylie T., Underwood K., Steptoe M., Theising B., Allen M., Bowers
 Y., Person B., Swaller T., Gibbons M., Pape D., Harvey N., Schurk
 R., Ralster E., Kohn S., Shin T., Jackson Y., Cardenas M., McCann
 R., Waterston R. and Willson R.

TITLE Public Soybean EST Project

JOURNAL Unpublished (1999)

COMMENT Contact: Shoemaker R./Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

This clone is available through: ResGen, Invitrogen Corp. 2130
 South Memorial Parkway Huntsville, AL 35801 For further information
 call: (800)-533-4363 or contact via email: c@resgen.com

High quality sequence stop: 405.

FEATURES

source

1..648
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl059-939"
 /clone.lib="Gm-cl059"
 /tissue_type="whole seedling, 2 week old, etiolated"
 /note="Vector: pBluescript II SK+; Site.1: EcoRI; Site.2:
 XhoI; The cDNA library was constructed from mRNA isolated
 from 2 week old etiolated whole seedlings of P1468916.
 Complementary DNA was synthesized from mRNA using a primer

consisting of a poly(dT) sequence with a XhoI restriction
 site. EcoRI adapters were ligated to the blunt-ended cDNA
 fragments followed by XhoI digestion. The cDNA fragments
 were directionally cloned into the EcoRI- XhoI restriction
 site of the plasmid vector. The ligated cDNA fragments
 were transformed into DH10B host cells (Gibco BRL). This
 library was constructed in the laboratory of Dr. Randy
 Shoemaker at Iowa state university."

BASE COUNT 173 a 132 c 161 g 181 t 1 others

alignment_scores:
 Quality: 1111.00 Length: 216
 Ratio: 5.241 Gaps: 0
 Percent Similarity: 98.148 Percent Identity: 94.444

alignment_block:
 US-09-701-572-2 x BG044933 ..

Align seg 1/1 to: BG044933 from: 1 to: 648

227 GATThrAsnAsnGlyValGlnIleTPAspAlaAlaArgCysLysIle 243
 1 GGAACAGCATGATGTAAGTTCAGATTGGATGCATCTGCATGCAGAGA 50
 243 sLeArgSerMetGluGlyHisArgLeuArgValGlyAlaLeuAlaTrpS 260
 51 GATAAGATCATGAGAGGCATCGGTACGGTGGCGCTGGCTTGA 100
 260 eArGThrLeuLeuSerSerGlyGlyArgAspLysAsnIleTyrGlnArg 276
 101 GTTCATCTCTTTGCTCTGGCGGAGGATAGATATTATCAAA 150
 277 AspIleArgThrGlnGluAspPheValSerLysLeuSerGlyHisLys 293
 151 GATATCCGTGCACAGAGATTTTGTACATAATGTCCAGGACAAATC 200
 293 rGluValCysGlyLeuLysTrpSerTyrAspAsnArgGluLeuAlaSerG 310
 201 AGAGGTTTGGAGCTGAGTGGCTTATGATTAACGCTGAGTGCATCTG 250
 310 lGlyLysAsnAspAsnLysLeuPheValTrpAsnGlnHisSerThrGln 326
 251 GAGGAATGACAAACAGATTTGTTGGAAATCAACACTCACTCAGCCT 300
 327 ValLeuLysTyrCysGlnHisThrAlaValLysAlaIleAlaTrpSe 343
 301 GTCTTAAGTACTGTGATGATGATGATGATGATGATGATGATGATG 350
 343 rProHisLeuHisGlyLeuLeuAlaSerGlyGlyThrAlaAspArgC 360
 351 TCTCATCTTCACGAGCTCTTGCATCTGGGGAGAGATCGACAGCAT 400
 360 ySLeArgPheTrpAsnThrThrAsnSerHisLeuSerCysMetAsp 376
 401 GCATACGTTTCGAAATACACCAACTCACTTAAGCTGATGAGAC 450
 377 ThrGlySerGlnValCysAsnLeuValTrpSerLysAsnValAsnGlu 393
 451 ACGGGAAGTCAGGTTTGCATCTCTGCTGCTCAANATGTCAATGAAC 500
 393 uValSerThrHisGlyTyrSerGlnAsnGlnIleLeuAlaTrpArgTyrP 410
 501 ACTAAGACACATGGCTATCCACAGACAGATTTGTTGGATGATACC 550
 410 rThrMetSerLysLeuAlaThrLeuThrGlyHisThrTyrArgValLeu 426
 551 CCTCATGTCAAAGTTGGCACTTACGGGATCAATCCATCCAGATCTCT 600
 427 TyrLeuAlaIleSerProAspGlyGlnThrIleValThrGlyAlaGly 442
 601 TATCTTGCATGTCTCCGATGGCAGACTATTGTACCTGGAGCTGGG 648

seq_name: gb_est2:BG444274

seq_documentation_block:

LOCUS BG444274 987 bp mRNA linear EST 15-MAR-2001

DEFINITION GA_Ea0023N11f Gossypium arboreum 7-10 dpa fiber library Gossypium arboreum cDNA clone GA_Ea0023N11f, mRNA sequence.

ACCESSION BG444274

VERSION BG444274.1 GI:13353926

KEYWORDS EST.

SOURCE Gossypium arboreum.

ORGANISM Gossypium arboreum

REFERENCE Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids II: Malvales: Malvaceae: Gossypium.

AUTHORS Wang, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry, D., Wood, T.C., Leslie, A. and Wilkins, T.A.

TITLE An integrated analysis of the genetics, development, and evolution of the cotton fiber

JOURNAL Unpublished (2000)

COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAAATGACATCTGACATAGGG
High quality sequence start: 40
High quality sequence stop: 800.

FEATURES

source

1..987
location/Qualifiers
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea0023N11f"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/note="Vector: PBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 258 a 203 c 256 g 265 t 5 others

ORIGIN

alignment_scores:

Quality: 1102.00 Length: 219
Ratio: 5.126 Gaps: 1
Percent Similarity: 98.174 Percent Identity: 94.064

alignment_block:

US-09-701-572-2 x BG444274 ..

Align seg 1/1 to: BG444274 from: 1 to: 987

254 VALGIALALEUALATRPSESRSESRLEUASERGLYARGAS 270
||| |||||:::|||||:::|||||:::|||||:::|||||
92 GIGAGGGCTTATCTTGAGCTCTTCTGTTATCTTCTGTAAGTCGTA 141
270 pYASnIleYrGlnArgAspIleArgThrGlnGluAspPheValSerL 287
|||||:::|||||:::|||||:::|||||:::|||||
142 TAAGACATCTTCAAGAGATATACGTCGACAGATGATTTTGTCTAGTA 191
287 yLSuSerGlyHsLysSerGluValCysGlyLeuLysTrpSerTrpAsp 303
|||||:::|||||:::|||||:::|||||:::|||||
192 AACTCTCTGTCACACAGTCAGAGCTTGTGGCTGTAAGTCCTATGAC 241
304 AsnArgGluLeuAlaSerGlyValAsnAspAluLysLeuPheValTrpAs 320
|||||:::|||||:::|||||:::|||||:::|||||
242 AACCGGAGCTGACATCTGCGAATGACACAACTTTTGTGTTGACA 291

320 nGlnHsSerThrGlnProValLeuLysTrpCysGlnHsThr AlaAla 336
|||||:::|||||:::|||||:::|||||:::|||||
292 TCACATCTCCACCCACACCTGTTCTTAATCTGTGACCATCTAGCTGCC 341
337 ValLysAlaIleAlaLarPseProHsLsGlyLeuLeuAlaSerG1 353
|||||:::|||||:::|||||:::|||||:::|||||
342 GTAAAGCAGTGTGTTGCTCCGACATCTGCATGCGCTTCTTCATCTGG 391
353 yGlyGlyThrAlaAspArgCysIleArgPheTrpAsnThrThrTrpAs 370
|||||:::|||||:::|||||:::|||||:::|||||
392 GGGGTGTACAGCTGATGATGATGATGATGATGATGATGATGATGATG 441
370 eHsLsSerCysMetAspThrGlySerGlnValCysAsnLeuValTrp 386
:::|||||:::|||||:::|||||:::|||||
442 CACACTGACCTGCATGACATGACATGACATGACATGACATGACATGAC 491
492 TCTAAGATGTTATGATGATGATGATGATGATGATGATGATGATGATG 541
403 nIleIleValTrpArgTrpProThrMetSerLysLeuAlaThrLeuThr 420
|||||:::|||||:::|||||:::|||||:::|||||
542 GATATGTTGTTGGATGATGATGATGATGATGATGATGATGATGATGAT 591
420 LysHsThrTrpArgValLeuTrpLeuAlaIleSerProAspGlyGlnTr 436
|||||:::|||||:::|||||:::|||||:::|||||
592 GACACACATACAGAGTCTTATCTTCAATCTCACCTGATGACAGACC 641
437 IleValThrGlyAlaGlyAspGluThrLeuArgPheTrpAsnValPhe 453
|||||:::|||||:::|||||:::|||||:::|||||
642 ATACTCTGAGCTGAGATGAGACGCTAAGCTTTGGAACTGTGCC 691
453 oSerProLysSerGlnAsnThrGlnSerGluIleGlyAlaLeuSerLeu 470
|||||:::|||||:::|||||:::|||||:::|||||
692 ATCCCTTAATCTCAGAACACTGACACTGAATAGAGCATCTCTTGG 741
470 LysArg 471
|||||
742 GAAGA 746

seq_name: gb_est1:AM030735

seq_documentation_block:

LOCUS AM030735 710 bp mRNA linear EST 18-MAY-2001

DEFINITION EST273390 tomato callus, TAMU Lycopersicon esculentum cDNA clone

ACCESSION CLEC25N23, mRNA sequence.

VERSION AM030735

KEYWORDS EST.

SOURCE Lycopersicon

ORGANISM Lycopersicon esculentum

REFERENCE Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Asteridae: euasterids I: Solanales; Solanaceae; Solanum; Lycopersicon.

AUTHORS 1 (bases 1 to 710)
Alcala, J., Vredalov, J., White, R., Matern, A.L., Vision, T., Holt, I.E., Liang, F., Dpton, J., Craven, M.B., Bowman, C.L., Ann, S., Rongling, C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.

TITLE Generation of ESTs from tomato callus tissue

JOURNAL Unpublished (1999)

COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

FEATURES

source

1..710
location/Qualifiers
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEC25N23"


```

/clone.lib="tomato callus, TAMU"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/notes="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; supplier: Giovanni laboratory; clec - cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST library"
BASE COUNT      209 a      142 c      149 g      209 t
ORIGIN

alignment_scores:
    Quality: 1001.00      Length: 193
    Ratio: 5.268          Gaps: 0
    Percent Similarity: 98.446      Percent Identity: 94.819

alignment_block:
US-09-701-572-2 x AM030735 ..

Align seg 1/1 to: AM030735 from: 1 to: 710

283 AspheValSerLysLeuSerGlyHisLysSerGluValCysGlyLeuY 299
|||||
3 GATTTCGACGATGAGTGAAGTGCATTAATCAAGAGGTTGGGCTCA 52
299 STPSETYRASPAsnArgLysLeuAlaSerGlyGlyAsnAspAsnLysL 316
|||||
53 ATGGCTTATGATACCGTGAATTAAGTTCAGGTGGAATGATATCGCG 102
316 EupheValTPraSngLHisSerThGlnProValLeuLysTyrCysGlu 332
|||||
103 TTTTGTATGGAACAACCATTCACACACCTGCTGAAATCTGTGAG 152
333 HisThAlaValAlaValAlaLeuAlaTPSerProHisLeuSngLys 349
|||||
153 CATACCTCTCGGTTAAGCCATTCATGCTCCGCCATCTCCATGGGCT 202
349 uLeuAlaSerGlyGlyThAlaAspArgCysIleArgPheTPraSng 366
|||||
203 TCTAGCATCCGGTGGAGGACACATGTCATGATTAATTTGTGAGAC 252
366 hThThTPraSngSerHisLeuSerCysMetAspThrGlySerGlnVal 382
|||||
253 CCACCACTAATACACATCTCAGTTGATGACACTGCGAGTCTGCTG 302
383 AsnLeuValTPSerLysAsnValAlaSngLLeuValSerThHisGly 399
|||||
303 AATCTGTGTCGTCGAAGATGTCATGATTAATGTCACACATGCTTA 352
399 rSerGlnAsnGlnIleLeuValTPraArgTyrProThMetSerLysLeu 416
|||||
353 CTCTCAAAATCAGATATAGTTGAGATATCCGACATGTTCAATGAT 402
416 lathLeuThPrgLHisThTyrArgValLeuTyrLeuAlaIleSerPro 432
|||||
403 CTACTCTGAGGTCATACATATAGAGTCTTATATCTCTAATATCTCA 452
433 AspGlyGlnThrIleValThrGlyAlaGlyAspGlyThrLeuArgPhe 449
|||||
453 GATGGACAGACATTTGCACTGAGACAGAGATGAACACTTCGATTCG 502
449 pAsnValPheProSerProLysSerGlnAsnThrGlnSerGluIleGly 466
|||||
503 GAATGTTTCTCCCTTCCCAAGTCAAGAACACCGAGACTGAATATGG 552
466 lalLeuSerLeuGlyArgThrThrIleArg 475
|||||
553 CATCTTCGTTTGTGAGACACTCAGATCAGG 581
seq_name: gb_est2:B1456001

```

```

seq_documentation_block:
LOCUS      B1456001      872 bp      mRNA      linear      EST 21-NOV-2001
DEFINITION 603170445F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:5250130 5',
            mRNA sequence.
ACCESSION  B1456001
VERSION    B1456001.1 GI:15246657
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  1 (bases 1 to 872)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgaps-remail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L14M1631 row: 1 column: 11
            High quality sequence stop: 834.
            Location/Qualifiers
                1..872
                /organism="Mus musculus"
                /strain="C57/B6"
                /db_xref="taxon:10090"
                /clone="IMAGE:5250130"
                /clone_lib="NCI_CGAP_Mam5"
                /tissue_type="tumor, gross tissue"
                /dev_stage="7 months"
                /lab_host="DH10B"
                /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
                Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                Library constructed by Life Technologies. Investigators
                providing samples: Lothar Hennighausen/Robin Humphreys,
                NIH"

BASE COUNT      174 a      256 c      268 g      174 t
ORIGIN

alignment_scores:
    Quality: 976.00      Length: 288
    Ratio: 3.920          Gaps: 6
    Percent Similarity: 86.458      Percent Identity: 68.750

alignment_block:
US-09-701-572-2 x B1456001 ..

Align seg 1/1 to: B1456001 from: 1 to: 872

157 TyrLysValLeuAspAlaProAlaLeuGlnAspAspPheTyrLeuAsnLe 173
|||||
2 TTCAAGGTGCTGGAGCCGCCAGAGCTTCAGGACGCTCTACCTCACTT 51
173 uValAspTPSerSerHisAsnValLeuAlaValGlyLeuGlnLysCysY 190
|||||
52 GTGGACCTGGTCTCCCTCAACGTCCTAGCGTGGGGCTGGGCACTGGCG 101
190 aTyrLeuTPraSnaLacYsserSerLysValThrLysLeuCysAspLeu 206
|||||
102 TGTACTGTGGAGTGCATGACACCGACGATGACCCGCTCTGGAGCTC 151
207 GlyValAsp...AspCysValCysSerValGlyThrPalaglnArgGlyTh 222
|||||
152 TCTGAGGAAGGAGGACTCAGTCACTTCGCTGGCTGAGCGGGGAA 201
222 rHisLeuAlaValGlyThrAsnAsnGlyLysValGlnIleTPraSnaLac 239

```

```

:  ::::::::::::::::::::::::::::  ::::::::::::::::::::::::::::
202 CTTGTCGCGAGTAGGTACACACAGGGCTTCGTGAGATCTGGGACGCTG 251
239 lAaRGcylsYsIleAaRGseRmetgluYhIaRgluAaRgYalglY 255
252 CTGCTGGGAGAAAGCTGTCCATGCTGGAGGGCCACACAGCAGATGGGG 301
256 AAlaAuaIatRPsSerSerleuSerSerleuSerSerleuSerSerleu 272
302 GCGCTGGCGCTGGAGATGACCGAGTTCATCTGTGACCGGACCGCAT 351
272 nIleYrGlnaRgAspIleAaRgT.....GlnGluAspPheValS 286
352 GATTCACAGCGGATATCCGACACACACCCCTGCATGACAG.....C 395
286 eRlySleuSerSerleuSerSerleuSerSerleuSerSerleuSerSer 302
396 GCGCGCTGACGGCCACCGCAGAAAGTGTGGCTTAAGTGGTCACA 445
303 AsPaSnARggluLeuAlaSerleuYglYsnaAsPaSnlySleuPheVal 319
446 GACCAACACCTGCTTCCTCGGGGGCAATGACAAAGCTGCTGCTG 495
319 pAsnGlnHISerThrGlnProValleuSerleuSerleuSerleuSer 336
496 GAACCACTACTCTAAGCCCTGTGACAGCATATACGAGACCTGGCAG 545
336 lAvallyAlaIleAlaTrPserProHISleuHISglYleuLeuAlaSer 352
546 CCGTGAAGGCTATGCTGCTGCCACACACACATGAGATGCTGCGCAT 595
353 GlYglYglYThrAlaSPaRgCylleAaRgPheTrPaSnThrThaS 369
596 GGGGGGACGGCTGACCGCTGACATCCATCTGGAACACTCTGACAG 645
369 nSerHISleuSerCysMetAsPThrGlySerleuAlaCysAsnleuVal 386
646 TCAGCAGACTGAGTGCATGACACAGGCTCACAAAGTGTCAACCTGGC 694
386 rPserlyAsnValAsnGluLeuValSerThrHISglYTrSerGlnAs 402
695 GGTCCAAAGCAGCCCAATGAGCTGGTGAAGCACATGCTACACACAGA 744
402 nGlnIleleVal.TrPaRgTrProThrMetSerlySleuAla.Thrle 418
745 CCAAGATCCCTGCTTGGAAAGTACCCGCTTACCGCAGGTGCCAAAGCT 794
418 uThrGlnHISrThrYrArgValleuTrleuAlaIleSerProAspGly 435
795 CACTGGGCATCATTTGCTGCTCTACTGCTGACATG.TCCCTGATGGG 843
435 lNthrIle 437
844 AGGCCATA 851
seq_name: gb_est2:B6760896
seq_documentation_block: 833 bp mRNA linear EST 15-MAY-2001
LOCUS B6760896
DEFINITION 602717081F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4840604 5',
mRNA sequence.
ACCESSION B6760896
VERSION B6760896.1 GI:14071549
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 833)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

```

Email: cga@phs-r@mail.nih.gov
 Tissue Procurement: ARCC/DCTD/DTF
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMU)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LMU at:
 http://image.lln.gov
 Plate: LCM1672 row: a column: 21
 High quality sequence stop: 765.
 Location/Qualifiers

FEATURES
 source
 1..833
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4840604"
 /clone_lib="NIH_MGC_49"
 /tissue_type="melanotic melanoma, high MDR (cell line)"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pORF7; Site_1: xhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCAAGAG(G). Size-selected >500bp for average insert size
 1.8kb. Library constructed by Ling Hong in the Laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH_MGC
 library."

BASE COUNT 159 a 262 c 277 g 135 t
 ORIGIN
 alignment_scores:
 Quality: 917.00 Length: 273
 Ratio: 3.869 Gaps: 6
 Percent Similarity: 86.813 Percent Identity: 67.399

alignment_block:
 US-09-701-572-2 x B6760896 ..

Align seg 1/1 to: B6760896 from: 1 to: 833

```

182 leuAlaValAlGlyLeuGlyAsnCySValTrleuTrPaSnAlaCySserSe 198
2 CTGACGCGTGGGGCTGAGCACCTGCGTGTACCTGTGAGTGGAGTCCCTGAC 51
198 rlyValThrlySleuCySAsPleuGlyValAsP...AsPcyValCySs 214
52 CCAAGTGAAGGGGCTGTGTGACCTCTCAGTGAAGGAGGACTCAGTCACT 101
214 eRValGlyTrPaAlaGlnARgglYhThrHISleuAlaValGlyTrAsPaSn 230
102 CCGTGGGCTGTGTGAGCGGGGGAACCTGGTGGCGGTGGGCACACACAG 151
231 GlYlys.ValGlnIleTrPaSpAlaAlaARgCySlySlySleAaRserM 247
152 GTGCTTGTGCGAGATCTGGGACCGACCGCAGGAGGAAACCTGTCACTGT 201
247 eTgluGlyHISARgLeuARgValAlGlyAlaLeuAlaTrPserSerSerleu 263
202 TGGAGGGCCACACAGCGCGCTCGGGGCGCTGGCTGGAATGCTGACAGAG 251
264 leuSerSerleuGlyARgAsPlyAsnIleYrGlnARgAspIleAaRgTh 280
252 CTGTCGTCCGGGAGCCGCGACCGATGATCTCGACAGAGACATCCGCAC 301
280 r.....GlnGluAspPheValSerlySleuSerGlyHISlySerg 294
302 CCGCCACCTGCAgTTCGGAG.....CGGCGGCTGACAGGCGACCGGAGG 345
294 luValCySglYleuLySTrPserTyRAsPaSnARggluLeuAlaSerGly 310
346 AGTGTGGGGCTCAAGTGTGTCACAGACACACCAAGCTCTCGCTGGGG 395

```

311 GLYAspAspAsnLysLeuPheValTrpAsnGlnHisSerThrGlnProVa 327
 |||||.....
 396 GCGAACGACAAACAGCTGCTGCTGCAATCACTGACCTGACCTGCTGCTGCT 445
 327 IleuLysTyrCysGlnHisThrAlaAlaValLysAlaIleAlaTrpSerP 344
 |||||.....
 446 GCGAGCTGACGAGCAGCTGCTGCTGCAAGGCTGCAAGCTGCTGCTGCTGCT 495
 344 roHisLeuHisGlyLeuLeuAlaSer.GlyGlyGlyThrAlaAspArgC 360
 |||||.....
 496 CACATGACGACGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 545
 360 sIleArgPheTrpAsnThrThrAsnSerHisLeuSerCysMetAspT 377
 |||||.....
 546 TATCCCTCTTCTGGAACACGCTGCTGCAACACACCTGCTGCTGCTGCTGCT 595
 377 hrgLysSerGlnValLysAsnLeuVal.TrpSerLysAsnValAsnGluLe 393
 |||||.....
 596 GCGGCTGCCAAGTGTCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 645
 393 uValSerThrHisGlyTyrSerGlnAsnGlnIleLeuValTrpArgTyrP 410
 |||||.....
 646 GGTGACGACGACGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 695
 410 roThrMetSerLysLeuAlaThrLeuThrGlyHisThrTyrArgValLeu 426
 |||||.....
 696 CCTCCCTGGAACGAGTGGCCAACTGACCGGCGACCTCCACCGCGCTG 745
 427 TyrLeuAlaIleSerProAspGlyGlnThrIleValThrGlyAlaGlyAs 443
 |||||.....
 746 TACCTGGCAATGTC.CCTGATGCGGAGGCGCATGCTGCTGCTGCTGCTGCT 794
 443 pglThrLeuArg 447
 |||||.....
 795 CGAGACCCCTGAGG 807

seq_name: gb_est2.B1652526

seq_documentation_block:

LOCUS B1652526 786 bp mRNA linear EST 12-SEP-2001

DEFINITION 603299431P1 NIH_CGAP_Mam3 Mus musculus cDNA IMAGE:5340103 5',

mRNA sequence.

ACCESSION B1652526

VERSION B1652526.1 GI:15566762

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 786)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cga@psh.remail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LAM11863 row: n column: 08

High quality sequence stop: 783.

FEATURES

source

1..786

location/Qualifiers

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone_image="5340103"

/clone_lib="NIH_CGAP_Mam3"

/tissue_type="tumor, gross tissue"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: NotI;

Site_2: Salt: Cloned unidirectionally. Primer: Oligo dr.
 Average insert 2 kb. Library constructed by Life
 Technologies, catalog #12017-018. Investigators providing
 samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference
 for transgenic model: Xu et al., Nature Genetics 22, 37-43
 (1999). Note: this is a NIH_CGAP Library."

BASE COUNT 170 a 231 c 233 g 152 t
 ORIGIN

alignment_scores:
 Quality: 902.50 Length: 251
 Ratio: 4.102 Gaps: 3
 Percent Similarity: 87.649 Percent Identity: 67.331

alignment_block:
 US-09-701-572-2 x B1652526

Align seg 1/1 to: B1652526 from: 1 to: 786

221 GLYThrHisLeuAlaValAlGlyThrAsnAsnGlyLysValGlnIleTrpAs 237
 |||||.....
 24 GGAACCTTGTCGCGAGTAGTACACACAAAGGCTTGGTGCAGATCTGGGA 73
 237 PALAlaArgCysLysLysIleArgSerMetGluClyHisArgLeuArgV 254
 |||||.....
 74 CGCTGCTCTGCGGAAGCTGTCATGCTGAGGGCCACACACACAGAG 123
 254 ALGlyAlaLeuAlaTrpSerSerLeuLeuSerSerGlyLysArgAsp 270
 |||||.....
 124 TGGGGGCGCTGCGCGAATGCTGACACAGTGTGTCATCTGTAAGCGGTGAC 173
 271 LysAsnIleTyrGlnArgAspIleArgThr.....GlnLysAspPh 284
 |||||.....
 174 GCGATGATCTTACACAGCGGATATCGCACACACCCCTGCTGCTGCTGCTGCT 221
 284 eValSerLysLeuSerGlyHisLysSerGlyValCysGlyLeuLysTrpS 301
 |||||.....
 222 ...CGGGGCTGCGAGGGCGCACCGGCAAGAGTGTGGCTTAAGTGTGT 267
 301 eTyrAspAsnArgGlyLeuAlaSerGlyGlyLysAspAsnLysLeuPhe 317
 |||||.....
 268 CCACAGACCACACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 317
 318 ValTrpAsnGlnHisSerThrGlnProValLeuLysTyrCysGlnHisTh 334
 |||||.....
 318 GTGTGGAACCACTAGCTGAAGCCCTGTGCAGCAGTATAGGAGACACT 367
 334 rAlaAlaValLysAlaIleAlaTrpSerProHisLeuHisGlyLeuAla 351
 |||||.....
 368 GCGACCGCTGAAGGCTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 417
 351 lAserGlyGlyThrAlaAspArgCysIleArgPheTrpAsnThrThr 367
 |||||.....
 418 CATCTGGTGGTGGCGACGGCTGACCCCTGCATTCGATTTGTGAACACTGTG 467
 368 ThrAsnSerHisLeuSerCysMetAspThrGlySerGlnValLysAsnLe 384
 |||||.....
 468 ACAGGTGACGACCTGACATGATGACACAGGCTCACACAGGTGTCACACT 517
 384 uValTrpSerLysAsnValAsnGluLeuValSerThrHisGlyTyrSerG 401
 |||||.....
 518 GCGCTGTGTCAAAGCGCCAAAGTGAAGTGTGTAAGCAACATGCTGCTGCTAC 567
 401 lAsnGlnIleIleValTrpArgTyrProHisMetSerLysLeuAlaThr 417
 |||||.....
 568 AGAACCAATCTCTCTGTGGAAGTACCGCTGCTTACCCAGTGGCCAG 617
 418 LeuThrGlnHisThrTyrArgValLeuLysLeuAlaIleSerProAspG 434
 |||||.....
 618 CTCACGTGCGCACATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 667
 434 yGlnThrIleValThrGlyAlaGlyAspGluThrLeuArgPheTrpAsnV 451

```

668 GGAGCGCCATAGTCCCGAGCTGAGATGAGACCCGTGAGTTCTGTGAATG 717
|||||
451 alphaProserProlySergInasThglusSerIulegylalaLeu 467
|||||
718 TCTTC.....AGCAAAACAACGCTCTAACAAAGAAATCTGTCTGCCTC 761
|||
468 Ser 468
:::
762 AAC 764

seq_name: gb_est2:BI904176

seq_documentation_block:
LOCUS      BI904176          913 bp    mRNA       linear   EST 16-OCT-2001
DEFINITION 603166804.F1 NCI_CGAP_Lu3 Mus musculus cDNA clone IMAGE:5255064 5'.
RNA sequence.
ACCESSION  BI904176
VERSION    BI904176.1  GI:16166369
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 913)
AUTHORS   NIH-MGC hlup://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: rcgabs@email.nih.gov
          Tissue Procurement: Gilbert Smith, Ph.D.
          CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
          Boraldo, Ph.D.
          CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
          DNA sequencing by: Incyte Genomics, Inc.
          Clone distribution: NCI-CGAP clone distribution information can be
          found through the I.M.A.G.E. Consortium/LNL.at:
          http://image.lnl.gov
          plate: LHAM1644 row: g column: 01
          High quality sequence start: 47
          High quality sequence stop: 901.
FEATURES             Location/Qualifiers
     source            1..913
                     /organism="Mus musculus"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone_image="5255064"
                     /clone_id="NCI_CGAP_Lu3"
                     /tissue_type="pooled lung tumors"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pRT30-Pac (Pharmacia) with a
modified polylinker. Site.1: NotI. Site.2: EcoRI. 1st
strand cDNA was prepared from mRNA obtained from pooled
lung tumors with a Not I - oligo(dT) primer [5'-
TGTTACCAATCTGAAAGTGAGCGCGCCCTCTGTTTTTTTTTTT 3'}.
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pRT30 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Boraldo."
BASE COUNT      200 a      266 c      271 g      175 t
ORIGIN
alignment_scores:
Quality:         896.50           Length:        252
Ratio:           4.094            Gaps:          3
Percent Similarity: 86.905       Percent Identity: 67.063

alignment_block:
US-09-701-572-2 x BI904176 ..

Align seg 1/1 to: BI904176 from: 1 to: 913
```

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 860)
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1AM12219 row: P column: 13
High quality sequence stop: 668.

FEATURES
source location/Qualifiers
1..860
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5534076"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-Sport6; Site_1: Not;
Site_2: Salt; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb.
281 c 280 g 138 t

BASE COUNT 161 a 281 c 280 g 138 t
ORIGIN

alignment_scores: Quality: 895.00 Length: 281
Ratio: 3.874 Gaps: 7
Percent similarity: 82.206 Percent identity: 64.769

alignment_block:
US-09-701-572-2 x BM462011 ..

Align seg 1/1 to: BM462011 from: 1 to: 860

```

125 SerMetHisSerLeuSerProPheMetAspAspPheValProGlyVal 141
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
30 TCTCCCTACTCCCTCTCTCCGTCAGCAACAGAGC.....CAGAGCT 73
141 IAsnHisSerProValLysAlaProArgLysValProArgSerProTyrL 158
: :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
74 GCTCCGTCGCCCGCGAACCACCCGCAAGATCTCCAAAGATCCCTTCA 123
155 yValLeuAspAlaProAlaLeuGlnAspAspPheTyrLeuAsnLeuVal 174
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
124 AGGTCTGTGACGCGCCGAGCTGACGAGCACTCTTACTCTCAATCTGTTG 173
175 AspTrpSerSerHisAsnValLeuAlaValGlyLeuGlyAsnGlyValTyr 191
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174 GACTGTGTCTCCCTCAATGTGTCTCAGCTGGGCTGAGGCACCTGCTGTA 223
191 rLeuTrpAsnAlaCysSerSerLysValThrLysLeuGlyAsnLeuGly 208
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
224 CCTGTGAGATGCTGTACACGACGAGTGTGAGCTGTGACCTCTCAG 273
208 aAsp...AspCysValCysSerValGlyTyrAlaGlnArgGlyThrHis 223
|||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
274 TGGAAAGGGAGCTCAGTGACCTCGTGGCTGGTCTGACGGGGGAACCTG 323
224 LeuAlaValGlyThrAsnAsnGlyLysValGlnIleTyrPaspAlaAla 240
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
324 GTGGCGGTGGGACACACAGAGGCTTGTGAGATCTGGAGACGAGCGCG 373

```

seq_name: gb_est1:A1895812

seq_documentation_block:

LOCUS A1895812 501 bp mRNA linear EST 18-MAY-2001
DEFINITION EST265255 tomato callus, TAMU Lycopersicon esculentum cDNA clone
CLC10M13, mRNA sequence.
ACCESSION A1895812
VERSION A1895812.1 GI:5601714
KEYWORDS EST.

SOURCE

tomato.
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; eunasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 501)
Alcala, J., Vredalov, J., White, R., Matern, A.L., Vision, T., Holt, I.E.,
Liang, F., Upton, J., Craven, M.B., Bowman, C.L., Ann, S., Roming,
C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

REFERENCE

AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES

source location/Qualifiers
1..501
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLC10M13"

```

240 gCysIysLysIleArgSerMetClnGlyHisArgLeuValGlyAla 257
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
374 AGGAAAGAGAGCTGTCCATGTGTGAGGGCCACAGCGACACGCTGGGGCGC 423
257 euAlaTrpSerSerLeuLeuSerSerGlyLysArgSpLysAsnLe 273
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
424 TGGCCCGAATGCTGAGCAGCTGTCTCGGAGCCGCGACCGCATGATC 473
274 TyrGlnArgAspIleArgThr.....GlnGluAspPheValSer 287
474 CTGCAGAGGAGACATCCGACACCCCGCACATGAGTGGAG.....CGGG 517
287 sLeuSerGlyHisLysSerGlyValCysGlyLeuLysTrpSerTyrAsp 304
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
518 GTCGACGGGCCACCGCAGAGGTGTGCGGCTCAAGTGTCCACAGACC 567
304 snArgGluLeuAlaSerGlyLysAsnAspLysLeuPheValTyrAsn 320
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
568 ACCAGCTCTCGCTCGGGGGGACACACACAGCTGCTGTGTGAT 617
321 GlnHisSerThrGlnProValLeuLysTyrCysGlnHisThrAlaAla 337
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
618 CACTCGAGCGCTGAGCCCGTGCAGAGTACAGCAGACACTGGCGCGCT 667
337 LysAlaIleAlaTrpSerProHisLeuHisGlyLeuLeuAlaSer 353
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
668 GAAGGCATGCTGCTGTCTCCCATCAGCAGCGGCTGTGCTGGCTGGGG 717
354 GlyGlyThrAlaAspArgCysIleArgPheTrpAsnThrThrAsnSe 370
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
718 GGGGGCACACTACCGCTGTATCCGCTTGTGAACACCTGACAGAGA 767
370 rHisLeuSerCysMetAspThrGlySerGlnValCysAsnLeuValTyr 386
: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
768 ACCACTGACGTGTATGACACAGCGGCTCCCAAGTGTCAATCTGGCTGG 817
387 SerLysAsnValAsnGluLeu..ValSerThrHisGly 398
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
818 TCCAAAGCAGCGCAACAGCTGGGTGAGCAGCGACGGG 855

```

```

/clone.lib="tomato callus, TAMU"
/tissue.type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/Note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Supplier: Giovannoni Laboratory; CLEC - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST Library"
BASE COUNT      145 a      102 c      113 g      141 t
ORIGIN

alignment_scores:
  Quality: 892.00      Length: 166
  Ratio: 5.373      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 97.590

alignment_block:
US-09-701-572-2 x AI895812 ..

Align seg 1/1 to: AI895812 from: 1 to: 501

283 AspPheValSerLysLeuSerGlnHisLysSerGluValCysGlyLeu 299
|||||
3 GATTTTGTCAAGTAAAGCTGCTCATTAATACAGAGTTTGCGCTCAA 52

299 STPSPSTYrAspAsnArgGluLeuAlaSerGlyGlyAsnAspAsnLys 316
|||||
53 ATGGCTTATGATACCGGTAAATTAAGTCTCAGGTGAAATGAAACGCG 102

316 eupheValTPraSngInHisSerThrGlnProValLeuLysTYrCysGlu 332
|||||
103 TTTTGTATGAGAACAAACATCAACAAACCTGCTGAAATACCTGTGAG 152

333 HisThrAlaAlaValLysAlaIleAlaTPSPSPProHisLeuHisGly 349
|||||
153 CATACGCTGCGGTAAAGCCATTCATGTCGCCCATCTCCATCGGCT 202

349 uLeuAlaSerGlyGlyThrAlaAspArgCysIleArgPheTPraST 366
|||||
203 TCTACGATCCGGTGGAGGACACGCTGATGATGATTAATCTGGAA 252

366 hThrThrAsnSerHisLeuSerCysMetAspThrGlySerGlnValCys 382
|||||
253 CCACACCTATACACATCTCAGTTGCATGACACGCGCAGTCTCT 302

383 AsnLeuValTPSPSerLysAsnValAsnGluLeuValSerThrHisGly 399
|||||
303 AATCTGTGTGTGCGAAGATGTCATTAATTAATGACACACATGCTTA 352

399 rSerGlnAsnGlnIleIleValTPraTYrProThrMetSerLysLeu 416
|||||
353 CTCTCAAAATCAGATAATAGTTGGAGATATCCGACAAATGCTTAAGAT 402

416 lATHrLeuThrGlyHisThrTYrArgValLeuTYrLeuAlaIleSerPro 432
|||||
403 CTACTCTGACAGGTACATATAGTCTTAATCTGCTATATCTCA 452

433 AspGlyGlnThrIleValThrGlyAlaGlyAspGluThrLeuArgPhe 448
|||||
453 GATGGACAGCAATGTCTCAGTGGACAGAGATGAACACTTCGATTC 500

seq_name: gb_est2:BI554814

seq_documentation_block:
LOCUS      BI554814      863 bp      mRNA      linear      EST 05-SEP-2001
DEFINITION 603236792F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5289670 5',
mRNA sequence.
ACCESSION  BI554814
VERSION    BI554814.1  GI:15442128
KEYWORDS   EST.

```

```

SOURCE      house mouse.
ORGANISM    Mus musculus.
REFERENCE   1 (bases 1 to 863)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgaabs-remail.nih.gov
             Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
             CDNA Library Preparation: Life Technologies, Inc.
             CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM1732 row: h column: 23
             High quality sequence stop: 855.
FEATURES
  source
    1..863
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /clone="IMAGE:5289670"
    /clone.lib="NIH_CGAP_Mam3"
    /tissue.type="tumor, gross tissue"
    /lab_host="DH10B"
    /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;
    Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
    Average insert 2 kb. Library constructed by Life
    Technologies, catalog #12017-018. Investigators providing
    samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference
    for transgenic model: Xu et al., Nature Genetics 22, 37-43
    (1999). Note: this is a NCI_CGAP Library."
BASE COUNT      191 a      236 c      257 g      159 t
ORIGIN

alignment_scores:
  Quality: 876.50      Length: 249
  Ratio: 4.039      Gaps: 5
  Percent Similarity: 87.149      Percent Identity: 67.871

alignment_block:
US-09-701-572-2 x BI554814 ..

Align seg 1/1 to: BI554814 from: 1 to: 863

226 ValGlyThrAsnAsnGlyLysValGlnIleTPraSPAlaAlaArgCys 242
|||||
3 GTAGGTACACACAAGGCTTGTCAGATCTGGGAGCGCTGCTGGGA 52

242 slyIleArgSerMetGluGlyHisArgLeuArgValGlyAlaLeuAla 259
|||||
53 GAACCTGTCATGCTGGAGGGCCACACAGCAGCGGTGGGCGCTGGCCT 102

259 rPSPSPSerLeuLeuSerSerGlyLysArgAspLysAsnIleTYrGln 275
|||||
103 GGATTCGTGACCAATCTGTCATCTGTGAGCCGTGACGCCACATCTC 152

276 ArgAspIleArgThr.....GlnGluAspPheValSerLysLeu 289
|||||
153 CGGATATCCGACACACACCCCTGCACTGACAG.....CGGCGCTGCA 196

289 rGlyHisLysSerGluValCysGlyLeuLysTPSPSTYrAsnArgC 306
|||||
197 GGGCCACCGGAGGAGTGTGTGCTTAAGTGTCCACAGACACACAGC 246

306 lueuAlaSerGlyLysAsnAspAsnLysLeuPheValTPraSngInHis 322
|||||
247 TGTCTGCTGCGGGGGGCAATGACAAACAAGCTGCTGTGTGAGCACTCT 296

323 SerThrGlnProValLeuLysTYrCysGlnHisThrAlaAlaValLys 339

```

```

297 AGCTAAGCCCTGTGACAGATATACGAGACACCTGGACCGGTGAAGC 346
339 ATLEATPPTSPROHISLEUHLISGLYLEUHLAASERGLYGLYT 356
347 TATTGCTGTGTCACACACACAGATGACTGCGCATGTGGGGGCA 396
356 hralaasparcysilearphetrpansnhrthrthrasnserhslau 372
397 CGGCTGACCGCTCATCTCTGGAACACCTGACAGCTGACGACAC 446
373 SerCyMetAspThrGlySerGlnValCysAsnLeuValTrpSerLys 389
447 CAGTGATTTGACACAGGCTCACAAGTGTCACTGCGCTGGCCAGA 496
389 nValAsnGluLeuValSerThrHisGlyTrpSerGlnAsnGlnLeu 406
497 CGGCANTGAGCTGTGAGACACATGCTACTACAGAACACATCTCG 546
406 alTrpArgTyrProThrMetSerLysLeuAlaThrLeuThrGlyHis 422
547 TGTGGAAGTACCGCTCCCTACGACAGTGGCCAACTGCTGCGCA 596
423 TyrArgValLeuThrLeuAlaIleSer.ProAspGlyGlnThrIle 439
597 TATGCTGTCTTACTGCTGACATGCTCCCTGATGGGAGGCCATG 646
439 hrcGlyAlaGlyAspGluThrLeuArgPheThrAsnValPhe.Pro 455
647 CCGGAGCTGAGATGAGACCTGAGCTTCTGGAATGCTTTCAGGCA 696
455 OlySerGlnAsnThrGluSerGluIleGlyAlaLeuSerLeu 697
697 ACCTCT.....ACAAGGAATCTGTGTGTGCTCAACCTC 733

seq_name: gb_est2:BE790506

seq_documentation_block:
LOCUS BE790506 888 bp mRNA linear EST 20-SEP-2000
DEFINITION 60158211P1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3936770 5',
mRNA sequence.
ACCESSION BE790506
VERSION BE790506.1 GI:10211704
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 888)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgsrbs-remail.nih.gov
Tissue Procurement: DCTD/BTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at: image.lnl.gov
Plate: LDCM/81 row: F column: 03
High quality sequence stop: 777.
FEATURES
source
1..888
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3936770"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally

```

```

BASE COUNT 166 a 276 c 307 g 139 t
ORIGIN
alignment_scores:
Quality: 868.00 Length: 270
Ratio: 3.910 Gaps: 9
Percent Similarity: 82.222 Percent Identity: 65.185
alignment_block:
US-09-701-572-2 x BE790506 ..
Align seg 1/1 to: BE790506 from: 1 to: 888
150 ArgLysValProArgSerProTyrLysValLeuAspAlaLeuG 166
13 CGCAAGATCTCCAGATGCCCTTCAAGGTGCTGGAGCGCCGAGCTGCA 62
166 nAspAspPheTyrLeuAsnLeuValAspTrpSerSerHisAsnValLeu 183
63 GGACACATTTACTCAATCTGTGTGACTGCTGCTCCATATGTCTCA 112
183 LalaGlyLeuGlyAsnCysValTyrLeuTrpAsnAlaCysSerLys 199
113 GCGTGGGCTAGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 162
200 ValThrLysLeuCysAspLeuGlyValAsp...AspCysValCysSerVa 215
163 GTGACGCGGCTCTGTGACCTCTGACGTGAGAGGAGCTGACGTGCTG 212
215 LglYTrpAlaGlnArgGlyThrHisLeuAlaValGlyThrAsnGlyL 232
213 GGGCGGTGTGTGAGCGGGGAACTGTGTGGTGGGCAACAGAGGCT 262
222 yValGlnIleTrpAspAlaIleArgCysLysLysIleArgSerMetGlu 248
263 TGTGTCAGATCTGGAGACCCAGCCAGGAGGAAAGACTGTCATGTGGAG 312
249 GlyHisArgLeuArgValGlyAlaLeuAlaTrpSerSerLeuLeuSe 265
313 GGCCACAGGAGGACCGGCTGGGGGCTGGCTGGAATGCTGAGAGCTGC 362
265 rSerGlyGlyArgAspLysAsnIleTyrGlnArgAspIleArgThr... 280
363 GTCCGGGAGCGCGACCCATGATCTCTGACAGAGGACATCGCGACCCGC 412
281 .....GlnGluAspPheValSerLysLeuSerGlnHisLysSerGluVal 295
413 CACTGCAGTCGGAG.....CGCGGCTGCAAGGCCACCGCAGAGAGTG 456
296 CysGlyLeuLysTyrSerTyrAspAsnArgGluLeuAlaSerGlyGlyAs 312
457 TCGCGGCTCAAGTGTGTCACAGACACAGACAGCTCGCTGGGGGGGCA 506
312 nAspAsnLysLeuPheValITrpAsnGlnHisSerThrGlnProValLeu 329
507 CGACAAACAGCTGCTGTGTGAAATCACTGACGCTGACGCCCTGGCAG 556
329 ySTyrCysGlnHisThrAlaValAlaValSerAlaIleAlaTrpSerProHis 345
557 AGTACACGAGAGACCTGGCGGCGTGAAGGCCATGCTGTGCTCCCAT 606
346 LeuHisGlyLeuLeuAlaSerGlyGlyGlyThrAlaAspArgCys...IleA 362
607 CAGACAGGCGGTGCTGCTGCGGGGGGCGGACAGCTGACCGCTGTAATCC 656
362 rPheTrpAsnThrThrAsnSerHisLeuSerCysMetAspThr...G 378

```

cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

```

|||||
657 GCTTCGGAACACGCTGCACAGACACACCTGACGGTTATGCACACGG 706
|||||
378 lYserGlnValcysAsnLeuValtTTPSerLysAsnValaIngluLeuVal 394
|||||
707 GCTCCCAAGTGTGCATCTGGCTGTCCACACGACCAACAGCTGGTG 756
|||||
395 SerThrHisGly.....TyrSerGlnAsnGlnIleValtTTPArgTy 409
|||||
757 AAG...CACGCGACGGTAACTCCACAGACAGAGA...TCTTGTGGAGATC 800
|||||
409 rPro 410
|||||
801 CCT 804

seq_name: gb_est2:BI153551
seq_documentation_block:
LOCUS BI153551 873 bp mRNA linear EST 05-JUL-2001
DEFINITION 60291713JF1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5067822 5',
mRNA sequence.
ACCESSION BI153551
VERSION BI153551.1 GI:14613552
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 873)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LRAM1182 row: m column: 07
High quality sequence start: 15
Location/Qualifiers
1. 873
/organism="Mus musculus"
/strain="C2EHC II (fetal)"
/db_xref="taxon:10090"
/clone="IMAGE:5067822"
/clone_id="NCI_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
stem cell origin."
/lab_host="DH10B"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dn.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 186 a 263 c 263 g 162 t
ORIGIN

```

```

alignment_scores:
Quality: 866.00 Length: 240
Ratio: 4.144 Gaps: 2
Percent Similarity: 87.083 Percent Identity: 68.333

```

```

alignment_block:
US-09-701-572-2 x BI153551

```

```

Align seg 1/1 to: BI153551 from: 1 to: 873

```

```

231 GlyValGlnIleTTPAspAlaAlaArgCysLysLysIleArgSerMe 247

```

```

|||||
29 GCTTCGTCGACGATCTGGACGCTGCTGGGGAAGAGTGCATGCT 78
|||||
247 tGluGlyHisArgLeuArgValGlyAlaLeuAlaTTPSerSerLeuL 264
|||||
79 GGAGGGCCACACACACAGATGGGGGGCGCTGGCTGGAAAGCTGACAGT 128
|||||
264 euseSerGlyLArgAspLysAsnIleTyrGlnArgAspIleArg..T 280
|||||
129 TGTATCTGTGTAGCGGTGACCGCATGATCTTACAGCGGATATCCGACA 178
|||||
280 hrgIngluAspPheValSerLysLeuSerGlyHisLysSerGluValcys 296
|||||
179 CCACCCCTGCAGTCAGAGCGGGCTGCAGGGCCACCGCAGAGATGTGT 228
|||||
297 GlyLeuLysTTPSerTyrAspAsnArgGluLeuAlaSerGlyLysAsn 313
|||||
229 GGCTTAAAGTGTCCACAGACACACAGCTGCTGCTCGGGGGCAATGA 278
|||||
313 PAsnLysLeuPheValTTPAsnGlnHisSerThrGlnProValLeuLysT 330
|||||
279 CACACAGCTGCTGCTGTGAAACCATCTAGTCTAAGCCCTGTGCAGCAGT 328
|||||
330 yTCysGluHisThrAlaValAlaValAlaIleAlaTTPSerProHisLeu 346
|||||
329 ATACGAGACACCTGCGAGCGGTGAAGCTATGCTGCTGCCACACAG 378
|||||
347 HisGlyLeuLeuAlaSerGlyGlyThrAlaAspArgGlyLeuArgP 363
|||||
379 CATGAGCTGCTGGCATCTGGTGGGCGACGGCTGACCGCTGCATTCGATT 428
|||||
363 eTTPAsnThrThrAsnSerHisLeuSerCysMetAspThrGlySerG 380
|||||
429 CTGGAACACCTGTACAGCTGACGACCATGATGATGACACAGGCTCAC 478
|||||
380 lValcysAsnLeuValtTTPSerLysAsnValaIngluLeuValSerTh 396
|||||
479 AAGGTCAACCTGGCTGTCCAAAGCAGCCCAATGAGCTGGAGACACA 528
|||||
397 HisGlyTyrSerGlnAsnGlnIleIleValTTPArgTyrProThrMetS 413
|||||
529 CATGGCTACTCAGACAGACAGATCCTGCTGTGAAGTACCCGCTTAC 578
|||||
413 lYsLeuAlaThrLeuThrGlyHisThrTyrArgValLeuTyrLeuAla 430
|||||
579 GCAGGTGGCCAGCTCAGTGCACCATATGTCGTCCTTACTCTGCGCA 628
|||||
430 lSerProAspGlyGlnThrIleValThrGlyAlaGlyAspGluThrLeu 446
|||||
629 TGTCCCTGTATGGGAGGCGCATATGTCACCGAGCTGAGATGAGACCTG 678
|||||
447 ArgPheThrAsnValPheProSerProLysSerGlnAsnThrIleSerG 463
|||||
679 AGGTTCGGAATGTCTTC.....AGCAAAACAGCCTCTCAAGAAATC 722
|||||
463 uileGlyAlaLeuSerLeu 469
|||||
723 TGTGCTGTGCTCAAGCTC 741

seq_name: gb_est2:BI904107
seq_documentation_block:
LOCUS BI904107 756 bp mRNA linear EST 16-OCT-2001
DEFINITION 60316673JF1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5254579 5',
mRNA sequence.
ACCESSION BI904107
VERSION BI904107.1 GI:16166293
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 756)

```


AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
CONTACT Robert Strausberg, Ph.D.
COMMENT Email: cgabbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LHAM1643 row: b column: 20
 High quality sequence start: 68
 High quality sequence stop: 756.
 Location/Qualifiers

FEATURES

source

1. 756
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:5254579"
 /clone_lib="NCI CGAP Lu33"
 /tissue_type="Pooled lung tumors"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: Lung; Vector: pRT3D-Pac (Pharmacia) with a
 modified polylinker; Site: 1: NotI; Site 2: EcoRI; 1st
 strand cDNA was prepared from mRNA obtained from pooled
 lung tumors with a Not I - oligo(dT) primer [5'.
 TGTACCAATCTGAAGGAGGAGGCGCCGCTCTTTTCTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not
 I and Eco RI sites of the modified pRT3D vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 149 a 216 c 237 g 152 t 2 others
ORIGIN

alignment_scores:
 Quality: 862.00 Length: 244
 Ratio: 4.124 Gaps: 3
 Percent Similarity: 85.656 Percent Identity: 69.262

alignment_block:

US-09-701-572-2 x B1904107 ..

Align seg 1/1 to: B1904107 from: 1 to: 756

```

159 ValLeuAspAlaProAlaLeuGlnAspAspPheTyrLeuAsnLeuValAs 175
|||||
32 GTCTGGACGCGCCGAGAGCTTCTTA CGATTCTACCTCACTGCTGGA 80
175 pTTPSerHisAsnValLeuAlaValGlyLeuGlyAsnValTyrL 192
|||||
81 CTGGCTCTCCCTCAACGCTGCTAGTGGGCTGGGACACCTGCTGTACC 130
192 eutTPsnaIacysSerSerLyValThrLysLeuGlyAsnLeuGlyVal 208
|||||
131 TGTGGAGTGCATGCACACGAGGTGACCCGCTGTGACCTCTCTGTA 180
209 Asp...AspCysValCysSerValGlyTPAlaGlnArgGlyThrHisLe 224
|||
181 GAAGGGGACTCACTGACTTCGCTGGCTGCTGACGCGGGGACTTGTGT 230
224 uAlaValGlyThrAsnAsnGlyLysValGlnIleTPAspAlaAlaArgC 241
|||||
231 CGCAGTAGGTACACACAGGCTCTGTCGACATCTGGGACGCTGCTG 280
241 ystLysLysIleArgSerMetGluGlyHisArgLeuArgValGlyAlaLeu 257
|||||
281 GGAAGAGGCTTCATGCTGTGAGGGCCACACAGACAGAGTGGGGCGCTG 330

```

```

258 AlaTPSerSerSerLeuLeuSerSerGlyGlyArgAspLysAsnIleTyr 274
|||||
331 GCCTGGAATGCTGACACAGTGTGATCTGTGACCCGTCGACCCGATGATCT 380
274 rGlnArgAspIleArgThr.....GlnGluAspPheValSerLysL 288
|||||
381 ACAGCGGATATCCGACACACACCCCTGCATGACAG.....CGCGGC 424
288 euSerGlyHisLysSerGluValCysGlyLeuLysTyrSerTyrAspAsn 304
|||||
425 TGCAGGCGCCACCGGACAGAGTGTGCTTAAGTGTGCTCCACAGACAC 474
305 ArgGluLeuAlaSerGlyLysAsnAspAsnLysLeuPheValTPAsnG 321
|||
475 CAGCTGCTTGCTCGGGGGCAATGACAAACAGCTGCTGCTGTGGAACCA 524
321 nHisSerThrGlnProValLeuLysTyrCysGluHisThrAlaAlaVal 338
|||
525 CTCTAGTCTAAGCCTGTGCACAGATATACGAGACCTGGCGAGCGTGA 574
338 yAlaAlaIleAlaTPSerProHisLeuHisGlyLeuLeuAlaSerGly 354
|||||
575 AGCTATTGCTGCTGCTGCCACACACAGCATGACTCTGCTGCTGTGTGT 624
355 GlyThrAlaAspArgCysIleArgPheTPAsnThrThrAsnSerHis 371
|||||
625 GGCACGGCTGACCGCTGCATCGATTCTGGAACACTGTGACAGTTCACGC 674
371 sLeuSerGlyMetAspThrGlySerGlyValCysAsnLeuValTPSerL 388
|||||
675 ACTGCAGTGCATTTGACACAGGCTCAAGTGTGCAACCTGGCGCTGTCCA 724
388 yAsnValAsnGluLeuValSerThrHisGly 398
|||||
725 AGCAGCCCAATGAGCTGTGTGAGACACATGCG 756

```

